THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 59, PART 3 1 May 1947 No. 333

THE ADIABATIC TEMPERATURE CHANGES ACCOMPANYING THE MAGNETIZATION OF COBALT IN LOW AND MODERATE FIELDS

BY L. F. BATES AND A. S. EDMONDSON,
University College, Nottingham

MS. received 6 *September* 1946

ABSTRACT. The new method devised for the measurement of the small thermal changes which are associated with the step-by-step changes in the magnetization of ferromagnetic materials in fields not exceeding a few hundred oersteds has been used in the study of annealed and unannealed cobalt in the form of stout wire. The observed changes are relatively large and in striking contrast to those observed with iron and nickel. An attempt is made to explain them on the basis of modern concepts in ferromagnetism.

§ 1. INTRODUCTION

EXTENSIVE investigations of the temperature changes which occur when a ferromagnetic substance is taken through an ordinary or so-called "technical" hysteresis cycle were made by Bates and Weston (1941) in the case of nickel and several nickel-iron alloys, and the results were described in a paper, hereafter referred to as Paper I, in which references to earlier work by other experimenters may be found. The investigations were extended to specimens of Armco iron by Bates and Healey (1943), described as Paper II. The present communication deals with work on cobalt. This metal, in the form of annealed specimens of electrolytic origin, was examined by Okamura (1936) who mounted bars of cobalt 12 cm. long alternately with bars of German silver in a cylindrical frame, and arranged a system of some 31 to 47 thermocouples in series by connecting the appropriate ends of these bars with wires of copper and constantan. Such an arrangement could not be regarded as entirely satisfactory, as is borne out by the fact that the thermal changes recorded experimentally were some 10 to 20 per cent greater than those calculated from the areas of the corresponding hysteresis cycles. Moreover, in presenting his results, Okamura divided the observed thermal changes into two parts, which he termed reversible and irreversible respectively, and we find it difficult to understand the argument upon which this division is based.

The cobalt used in our work was kindly supplied by Messrs. Brandhurst & Co. Ltd. in the form of No. 12 S.W.G. wire. Its composition was : Co 98·40, Ni 0·43, Fe 0·13, CaO 0·23, Mn 0·08, C 0·19, Zn 0·01, Mg 0·11, SiO_2 0·14, S 0·02 per cent : the loss observed on heating in hydrogen was 0·24 per cent. The metal was originally cast into small ingots which were then cogged and rolled into bars from which the wire was drawn. Measurements were made on the material in the hard-drawn state, exactly as supplied, and also on wires which were annealed by heating them at 700° c. in an evacuated quartz tube for 60 minutes and thereafter allowing them to cool slowly. This annealing process is generally held to produce re-crystallization without undue increase in grain size.

A 40-cm. length of cobalt wire was mounted along the axis of the vertical water-cooled solenoid, as in Papers I and II. Adiabatic temperature changes of the wire were measured by means of twenty copper-constantan thermocouples. The "hot" junction of each couple was kept in moderately loose contact with the wire, while the "cold" junction was very close to, but thermally insulated from, the wire, except for conduction along the material of the couple. Moderately loose contact meant that the specimen was in no wise strained or prevented from changing its dimensions freely due to changes in magnetization. Each couple was joined to its own separate primary winding of insulated low-resistance copper wire wound upon a section of a mu-metal spiral core. A low-resistance secondary coil of many turns was wound upon this core and connected to a specially designed fluxmeter of high sensitivity. Electrical insulation between the "hot" junctions and the cobalt wire was not necessary but, in order to avoid instability of the fluxmeter zero, it was necessary to earth the specimen and all portions of the surrounding apparatus and, in addition, one of the leads to the moving coil of the fluxmeter was earthed.

When the temperature of the wire was rapidly changed by a small quantity ΔT, a ballistic deflection of the fluxmeter strictly proportional to ΔT took place. The whole system was normally calibrated in earlier work by suddenly applying a longitudinal force of F dynes to the wire, so causing an adiabatic fall in temperature ΔT_1 given by

$$\Delta T_1 = \frac{-\alpha TF}{J\rho SA},$$

where α is the coefficient of linear expansion of the cobalt, T its absolute temperature, J the mechanical equivalent of heat, while ρ, S and A are respectively the density, specific heat and area of cross-section of the wire. As the energy in ergs required to change the temperature of 1 c.c. of the wire by ΔT_1 is $J\rho S \,.\, \Delta T_1 = \alpha TF/A$, the density and specific heat need not be known in order to express the experimental results in the most convenient way.

Professor W. Wilson, in private conversation, has kindly pointed out to us that this method of calibration is based on the assumption that the thermodynamic conditions for a reversible change are satisfied, and these, in particular, require that the load should be applied slowly enough to ensure only very slight departure from equilibrium of the system at any time during the change. We think that our method of applying the load caused this requirement to be satisfied. It would appear, however, that equilibrium considerations have received little attention in dealing with the problems of the hysteresis cycle under alternating field conditions, as, for example, in considering the effect of changing the frequency on the rate of generation of heat per cycle for a given maximum intensity of an alternating magnetic field. Unfortunately, the loading method of calibration failed because the wire was so thin that it was impossible to mount it without bending. Hence, a sudden application of a longitudinal force produced irregular bending of the wire and gave fluxmeter deflections which were not strictly proportional to the force. We therefore assumed that in all cases Warburg's law was accurately obeyed, i.e. that the total heat liberated in the wire when it was taken through a complete hysteresis cycle was exactly equal to $\oint H\,dI$. This is only correct when eddy-current effects are negligible. Fortunately, towards the end of our measurements we were able to use a check method of calibration, devised by Mr. E. G. Harrison, which is based on the heating effect of a low-frequency alternating current passed through the wire for a known short interval of time, and which proved the above procedure to be sound. In any case, the fact that the values of the sensitivity of the system, as found from the data for the different cycles, assuming Warburg's law to hold, were in good agreement with one another showed that the method was reliable.

The necessary magnetic measurements were made by the ballistic method of Paper I. As the ratio of length to diameter of the specimen was so great, the value of the demagnetization coefficient was so small that it was not required to a high degree of accuracy, and it was therefore taken to be that for an ellipsoid of revolution with the appropriate dimensional ratio.

§ 2. EXPERIMENTAL DETAILS

The main sources of error found with this method were discussed fully in Paper I, and the same steps were taken to avoid errors due to zero drift in the fluxmeter, inadequate thermal insulation, eddy currents in the specimen and the effects of stray fields from solenoid and specimen upon the mu-metal core and upon the thermocouple leads. In particular, a 2-henry choke was connected in the solenoid circuit to reduce the rate of change of magnetization in the specimen and, consequently, the magnitude of the eddy-current heating. The latter was proved to be unimportant by the fact that the total heat generated in describing a given closed hysteresis cycle did not depend upon the number of steps or field changes in which it was done.

The fluxmeter itself was used under approximately the same sensitivity conditions as in Paper I, but calibration showed the overall sensitivity of the thermocouple-fluxmeter system to be disappointingly low, viz. about one-third that in Paper I. This was mainly because the "hot" junction contacts had been designed for specimens of larger diameter, with which greater areas of contact were possible. Attempts to improve the overall sensitivity by reducing the size of the junctions and altering the mode of attachment gave very little result.

The troublesome induction effect of unknown origin was present in magnitude rather greater than in Paper I but much less than in Paper II, and it persisted in spite of great care in arranging the thermocouple leads etc., as symmetrically as possible. As in Paper II, its effects were compensated by adjusting the compensating coil by trial prior to the recording of the data for a chosen step in the magnetizing current.

§ 3. EXPERIMENTAL RESULTS

The results for unannealed cobalt are given in figures 1 to 5. The first shows the three main hysteresis cycles used in the work. It is clear that measurements with much higher fields would have been very informative, as we barely attained even "technical" saturation, but these would have required a new solenoid of many more turns and a more complicated cooling system. Following the procedure of Papers I and II, the heat changes $\Sigma\, dQ$ are denoted by Q; these were recorded and summed, in the cases of figures 3, 4 and 5 *a*, as the effective soleniod field was

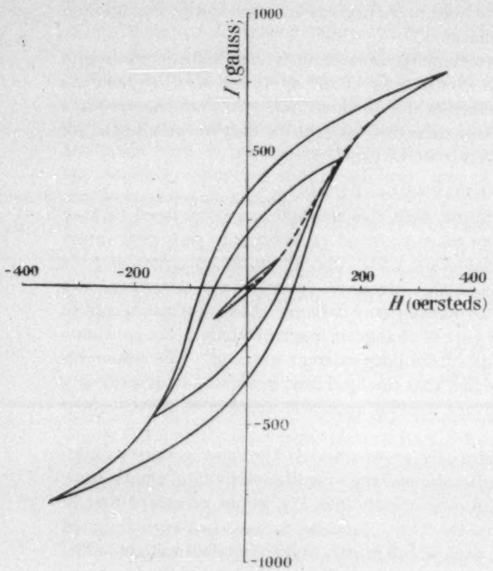


Figure 1. Hysteresis cycles for unannealed cobalt.

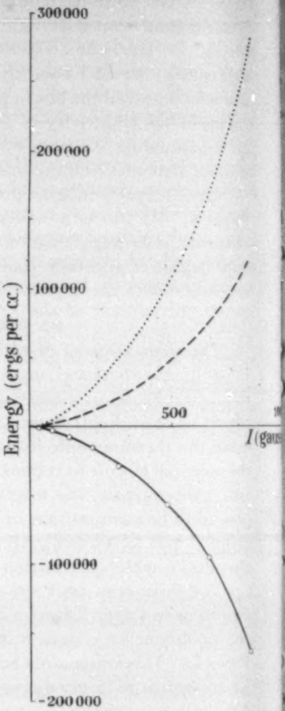


Figure 2. Virgin curves for unannealed co

—— Q, I; - - - - $\int H\, dI, I$;

...... $\int H\, dI - Q, I$.

changed step by step from the stated maximum value, $-H_m$, to an equal maximum, $+H_m$, in the opposite sense. The values of Q are plotted as a function of the observed intensity of magnetization of the specimen, and, in order to economize in graph space, the values for this half-cycle only are plotted. The changes which occurred in the half-cycle from $+H_m$ to $-H_m$ would give the graph obtained by rotating the existing Q curve about the axis of ordinates and displacing it vertically

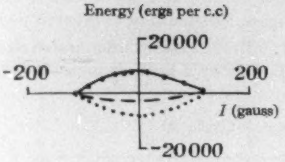


Figure 3. Unannealed cobalt. Cycle A. Maximum field 59 oersteds.

—— Q, I ; ---- $\int H\,dI, I$;

...... $\int H\,dI - Q, I$.

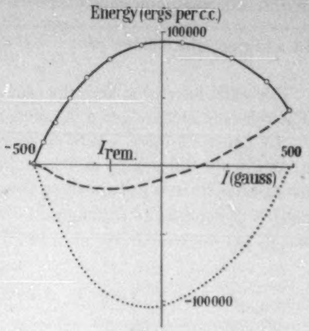


Figure 4. Unannealed cobalt. Cycle B. Maximum field 175 oersteds.

—— Q, I ; ---- $\int H\,dI, I$;

...... $\int H\,dI - Q, I$.

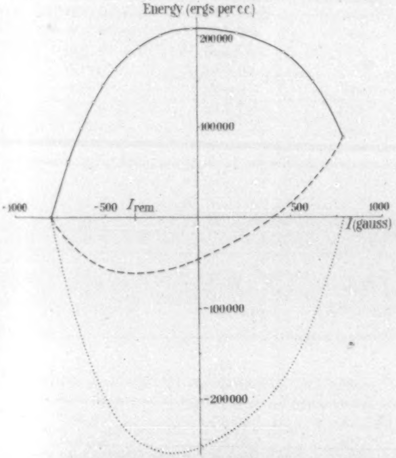


Figure 5*a*. Unannealed cobalt. Cycle C. Maximum field 351 oersteds.

—— Q, I ; ---- $\int H\,dI, I$;

...... $\int H\,dI - Q, I$.

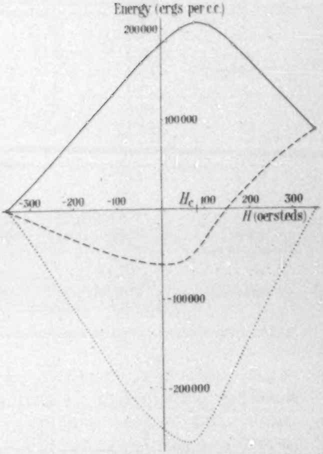


Figure 5*b*. Unannealed cobalt. Cycle C. Maximum field 351 oersteds.

—— Q, H ; ---- $\int H\,dI, H$;

...... $\int H\,dI - Q, H$.

until its starting point coincided with the point at which the existing curve ends. In addition, the values of $\int_{-H_m}^{H} H\,dI$ and $\int_{-H_m}^{H} H\,dI - Q$ are plotted against I.

The scales have been kept the same, wherever possible, to facilitate comparison. The relevant data are given in tables 1 to 4.

In figure 5 *b* these several quantities have been plotted against H in order to bring out the differences in behaviour of unannealed iron and nickel on the one hand and of unannealed cobalt on the other. Figure 5 *b* should be compared with figure 5 of Paper I and with figure 1 *b* of Paper II. In the cases of unannealed iron and nickel there is always an initial cooling followed by a heating such that the

Table 1. Unannealed cobalt—Cycle A

Step	H (oersteds)	I (gauss)	$\int H\,dI$ (ergs/c.c.)	Q (ergs/c.c.)	E (ergs/c.c.)
0	−59·1	−114·0	0	0	0
1	−45·3	−91·7	−1,130	+2,690	−3,820
2	−29·6	−66·7	−2,060	+5,220	−7,280
3	−15·0	−41·1	−2,640	+7,040	−9,680
4	−6·0	−25·3	−2,800	+7,200	−10,000
5	−0·5	−15·3	−2,840	+7,360	−10,200
6	+0·5	−13·6	−2,840	+7,360	−10,200
7	+15·1	+14·6	−2,620	+6,760	−9,400
8	+29·7	+46·1	−1,910	+5,600	−7,510
9	+45·3	+81·3	−580	+3,740	−4,320
10	+59·1	+114·0	+990	+990	0

Table 2. Unannealed cobalt—Cycle B

Step	H (oersteds)	I (gauss)	$\int H\,dI$ (ergs/c.c.)	Q (ergs/c.c.)	E (ergs/c.c.)
0	−175·0	−484	0	0	0
1	−143·0	−451	−1,500	+16,400	−17,900
2	−109·0	−407	−8,060	+32,700	−40,800
3	−73·6	−351	−12,200	+49,200	−61,400
4	−37·7	−286	−15,900	+64,800	−80,700
5	−0·5	−202	−17,500	+78,500	−96,000
6	+1·7	−197	−17,500	+78,500	−96,000
7	+38·8	−82	−15,100	+89,100	−104,200
8	+74·5	+76	+5,980	+91,600	−97,600
9	+110·0	+266	+11,500	+80,500	−69,600
10	+144·0	+396	+27,900	+63,400	−35,500
11	+175·0	+484	+42,000	+42,000	0

Table 3. Unannealed cobalt—Cycle C

Step	H (oersteds)	I (gauss)	$\int HdI$ (ergs/c.c.)	Q (ergs/c.c.)	E (ergs/c.c.)
0	−351·0	−790	0	0	0
1	−231·0	−699	−26,800	+ 61,300	− 88,100
2	−166·0	−628	−40,500	+ 98,000	−138,000
3	−111·0	−557	−50,600	+127,000	−178,000
4	− 76·5	−503	−55,800	+145 000	−201,000
5	− 43·2	−438	−59,600	+163,000	−223,000
6	− 10·5	−368	−61,600	+178,000	−240,000
7	+ 12·5	−307	−61,600	+187,000	−249,000
8	+ 45·1	−197	−58,400	+200,000	−258,000
9	+ 78·1	− 25	−47,500	+207,000	−254,000
10	+112·0	+218	−24,300	+203,000	−227,000
11	+166·0	+471	+10,500	+182,000	−172,000
12	+231·0	+631	+42,000	+150,000	−108,000
13	+351·0	+790	+87,600	+ 87,600	0

Table 4. Unannealed cobalt—Virgin curve

Step	H (oersteds)	I (gauss)	$\int HdI$ (ergs/c.c.)	Q (ergs/c.c.)	E (ergs/c.c.)
0	0·0	0·0	0	0	0
1	+ 0·4	+ 0·6	0	0	0
2	+ 22·9	+ 36·4	+ 450	− 800	+ 1,250
3	+ 45·1	+ 82·7	+ 2,040	− 2,700	+ 4,740
4	+ 67·6	+138·0	+ 5,180	− 7,800	+ 13,000
5	+ 90·8	+220·0	+ 11,700	− 15,700	+ 27,400
6	+115·0	+310·0	+ 21,100	− 29,000	+ 50,100
7	+168·0	+485·0	+ 45,700	− 57,000	+103,000
8	+232·0	+634·0	+ 75,400	− 96,000	+171,000
9	+351·0	+795·0	+121,000	−165,000	+286,000

maximum fall of temperature occurs in the interval of field change from $-H_m$ to $-H_c$. With cobalt there is always an initial warming, such that the maximum rise of temperature takes place in the field change from $-H_m$ to $+H_c$. Here, $+H_c$ means the coercive field actually required to reduce the magnetization to zero; heating is observed in all cases, of course, as the field is changed from $-H_c$ to $+H_c$, in accord with the view that, between these field limits, changes in magnetization take place almost entirely by irreversible 180-degree reversals in the domains.

The results for annealed cobalt are given in figures 6 to 10; the relevant data for figure 10 are given in tables 5 to 8. Incidentally, the results of figure 9 were

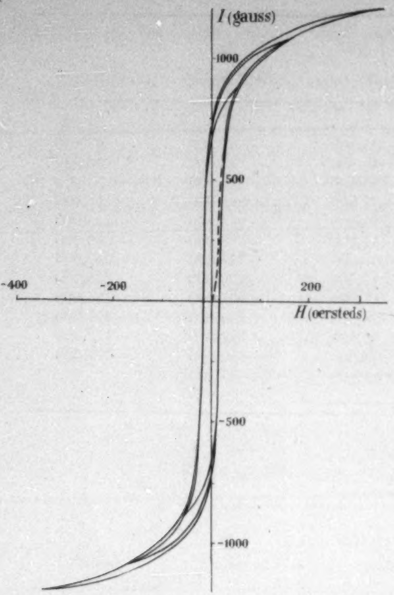


Figure 6. Hysteresis cycles for annealed cobalt.

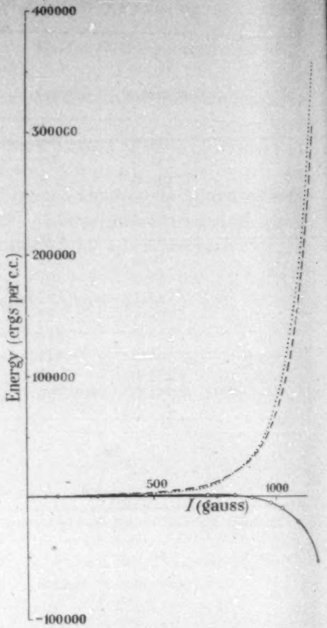


Figure 7. Virgin curves for annealed cobalt.

—— Q, I ; - - - - $\int H\,dI, I$;

...... $\int H\,dI - Q, I$.

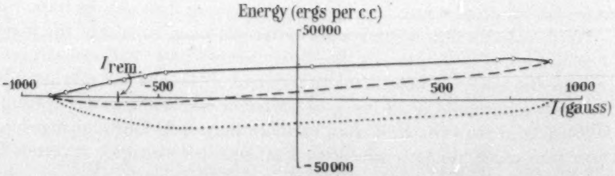


Figure 8. Annealed cobalt. Cycle A. Maximum field 57 oersteds.

—— Q, I ; - - - - $\int H\,dI, I$; ...... $\int H\,dI - Q, I$.

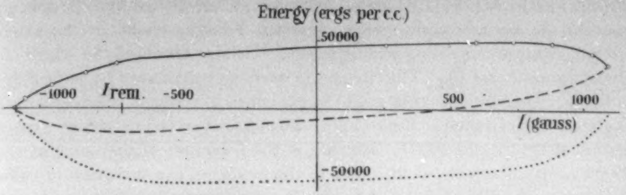


Figure 9. Annealed cobalt. Cycle B. Maximum field 173 oersteds.

——— Q, I ; - - - - $\int H\,dI, I$; . . . . . . $\int H\,dI - Q, I$.

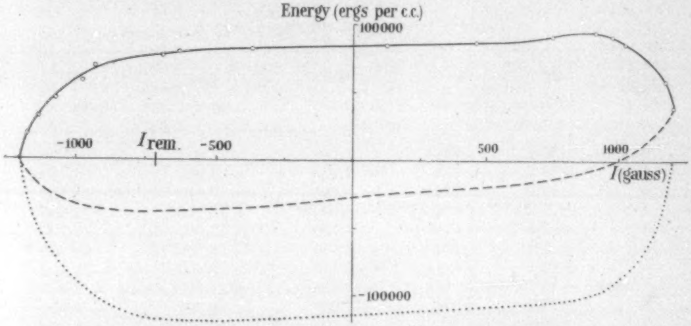


Figure 10. Annealed cobalt. Cycle C. Maximum field 350 oersteds.

——— Q, I ; - - - - $\int H dI, I$; . . . . . . $\int H dI - Q, I$.

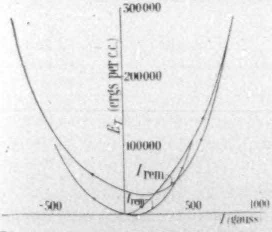


Figure 11 *a*. Curves of E_T against I for unannealed cobalt.

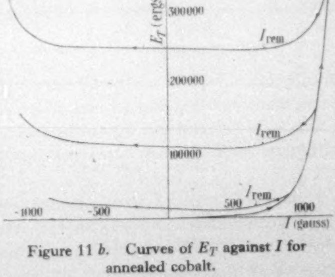


Figure 11 *b*. Curves of E_T against I for annealed cobalt.

fully confirmed by Mr. E. G. Harrison, who completely dismantled the specimen, overhauled the thermocouple system, inserted a larger choke in the solenoid circuit and changed its cooling arrangements. He also obtained the whole of the results of figures 7 and 10. The changes in coercivity produced by annealing are remarkable. Qualitatively, they might be explained in terms of the magnetization of single crystals of cobalt. Kaya (1928) showed that in cobalt, which possesses hexagonal structure, the (0001) direction is the direction of easy magnetization while the (1010) direction, in the plane of the hexagon, is a direction of difficult magnetization, requiring fields of some ten thousand oersteds to produce technical saturation. If in the hard-drawn state many of the crystals in the polycrystalline wire tended to set with their directions of easy magnetization perpendicular to the axis of the wire, it would be difficult to produce saturation parallel to that axis. If, however, the effect of annealing is to reduce the amount of internal strain in the

Table 5. Annealed cobalt—Cycle A

Step	H (oersteds)	I (gauss)	$\int H\,dI$ (ergs/c.c.)	Q (ergs/c.c.)	E (ergs/c.c.)
0	−56·5	−891	0	0	0
1	−34·1	−838	− 2,420	+ 3,020	− 5,440
2	−14·1	−758	− 4,160	+ 7,340	−11,500
3	− 3·2	−678	− 4,690	+10,900	−15,600
4	+ 2·0	−625	− 4,720	+13,000	−17,700
5	+ 7·0	−550	− 4,440	+15,500	−19,900
6	+10·1	−473	− 3,530	+18,400	−21,900
7	+16·2	+ 52	+ 6,650	+24,100	−17,400
8	+30·0	+704	+19,900	+29,200	− 9,300
9	+34·3	+760	+21,800	+30,000	− 8,200
10	+56·5	+891	+27,600	+27,600	0

Table 6. Annealed cobalt—Cycle B

Step	H (oersteds)	I (gauss)	$\int H\,dI$ (ergs/c.c.)	Q (ergs/c.c.)	E (ergs/c.c.)
0	−173·0	−1,090	0	0	0
1	−116·0	−1,050	− 5,210	+ 7,300	−12,500
2	− 53·9	− 956	−12,600	+17,200	−29,800
3	− 17·7	− 835	−16,700	+25,900	−42,600
4	− 2·2	− 727	−17,700	+32,800	−50,500
5	+ 13·4	− 416	−15,200	+39,200	−54,400
6	+ 19·4	+ 255	− 4,260	+46,600	−50,900
7	+ 26·0	+ 593	+ 3,040	+49,600	−46,600
8	+ 34·0	+ 733	+ 7,170	+50,200	−43,000
9	+ 54·1	+ 882	+13,600	+49,000	−35,400
10	+116·0	+1,030	+25,800	+41,500	−15,700
11	+173·0	+1,090	+33,300	+33,300	0

Table 7. Annealed cobalt—Cycle C

Step	H (oersteds)	I (gauss)	$\int H\,dI$ (ergs/c.c.)	Q (ergs/c.c.)	E (ergs/c.c.)
0	−350·0	−1,120	0	0	0
1	−252·0	−1,170	− 9,970	+18.900	− 28,900
2	−176·0	−1,130	−19,200	+31,800	− 51,000
3	−116·0	−1,070	−26,700	+44,800	− 71,500
4	− 54·7	− 976	−34,400	+58,000	− 92,400
5	− 37·5	− 930	−36,400	+68,500	−105,000
6	+ 2·3	− 692	−38,800	+76,200	−115,000
7	+ 6·8	− 632	−38,600	+79,200	−118,000
8	+ 13·7	− 371	−36,100	+81,800	−120,000
9	+ 18·1	+ 122	−24,700	+84,400	−109,000
10	+ 23·0	+ 458	−20,300	+86,600	−107,000
11	+ 36·2	+ 746	−14,200	+91,400	−106,000
12	+ 65·9	+ 909	− 6,220	+94,300	−100,000
13	+ 95·5	+ 973	− 120	+91,700	− 91,800
14	+116·0	+1,020	+ 3,260	+86,100	− 82,800
15	+265·0	+1,170	+29,300	+58,600	− 29,300
16	+350·0	+1,200	+38,700	+38,700	0

Table 8. Annealed cobalt—Virgin curve

Step	H (oersteds)	I (gauss)	$\int H\,dI$ (ergs/c.c.)	Q (ergs/c.c.)	E (ergs/c.c.)
0	0	0	0	0	0
1	+ 11·2	+ 121	+ 506	+ 377	+ 129
2	+ 15·3	+ 279	+ 1,270	+ 659	+ 611
3	+ 20·3	+ 468	+ 3,140	+ 800	+ 2,340
4	+ 32·0	+ 718	+ 10,200	+ 942	+ 9,290
5	+ 45·4	+ 824	+ 20,600	+ 612	+ 20,000
6	+ 58·5	+ 893	+ 31,800	− 1,080	+ 32,900
7	+ 85·3	+ 977	+ 57,100	− 5,220	+ 62,300
8	+ 114·0	+1,030	+ 85,400	−10,000	+ 95,400
9	+ 174·0	+1,100	+149,000	−21,000	+170,000
10	+ 349·0	+1,200	+307,000	−53,600	+361,000

wire and also permit recrystallization with the directions of easy magnetization more favourably directed, there would result an increase in remanence and a decrease in coercivity. However, from what follows, it appears that the result of annealing is to produce big changes in crystal structure instead of mere orientation effects.

The arched shape of the Q, I curves appears to be a distinctive feature of unannealed cobalt. If we compare figures 3, 4 and 5 *a* above with figures 13 and 14 of Paper I for annealed nickel, some slight resemblance is seen; it is somewhat intensified by annealing the cobalt, when the arches become much flattened

Indeed, the graphs of figures 8 to 10 between $-I_r$ and $+I_r$, i.e. between the two extreme values of the retentivity, are linear within the limits of experimental error. The magnitudes of the heat changes are considerably greater than those observed in corresponding cycles for iron and nickel.

In table 9 an attempt has been made to summarize the more distinctive features of the Q, I curves for the three ferromagnetic metals, and we see that, on the whole, cobalt behaves in a manner contrary to that of iron and nickel.

Table 9

	Field change	Heat changes in closed hysteresis cycles		
		Iron	Nickel	Cobalt
Unannealed metals	$-H_m$ to $-H_c$	Cooling	Cooling	Heating
	$-H_c$ to $+H_c$	Heating	Heating	Heating
	$+H_c$ to $+H_m$	Heating	Heating	Cooling
Annealed metals	$-H_m$ to $-H_c$	Cooling	Complicated	Heating
	$-H_c$ to $+H_c$	Cooling	Cooling	Heating
	$+4H_c$* to $+H_m$	Heating	Heating	Cooling

* $+4H_c$ is usually taken as a reasonable value of H at which technical saturation may be assumed complete; here, it is given merely as an indication that complicated phenomena around $H=+H_c$ are excluded.

Unfortunately, Bates and Healey were unable to obtain the virgin Q, I curve for iron; arrangements are now being made to do so, but success is doubtful owing to the special difficulties peculiar to iron. Consequently, only the results for nickel may be compared with those for cobalt. Noting the change of abscissae, figure 2 above may be compared with figure 6 of Paper I, showing that unannealed virgin cobalt cools as it is magnetized while unannealed nickel warms, and that there is in both cases the suggestion of a "knee" in the Q, H curve (not reproduced here).

In addition to the figures published in Paper I, Bates and Weston (Weston, 1940) made many measurements of the Q, I curves for annealed virgin nickel, finding that such curves rise very sharply from the origin to reach a maximum at a magnetization equal to approximately 70 to 80 per cent of the saturation value, when they drop steeply to cross the axis of abscissae to the side where cooling is represented. The contrast with figure 7 above is striking. Mr. E. G. Harrison has extended the range of the latter curve and finds that strong cooling is still shown as even higher values of I are reached. The question therefore arises as to the stage in magnetization at which the cobalt and the nickel curves turn upwards, for we know that every ferromagnetic metal exhibits the normal magnetocaloric effect, namely a heating directly proportional to the intensity of the strong magnetic field in which it is placed suddenly. As a matter of fact, Bates and Weston found that one specimen of nickel, specimen 1A, showed an upward turn at $H=100$ oersteds and $I=420$ gauss approximately. Incidentally, as this specimen of nickel was subjected to increasing longitudinal stress, the Q, I curve moved to lie in its entirety above the I axis, after which, with increasing tension, the curve approached closer and closer to that axis.

§4. DISCUSSION OF RESULTS

The curves in the preceding section have been plotted in the same manner as those of Papers I and II in order to facilitate comparison, but much is gained, for example, by plotting all the Q, I curves for annealed cobalt on the same graph. One then appreciates the flat initial portion of the virgin curve and the low arches of the succeeding half-cycle curves. In like manner, in the case of unannealed cobalt, one sees the steep descent of the virgin curve and the strongly arched curves of the succeeding half-cycles. Similarly the $\int H\,dI, I$ curves can be treated to emphasize the great differences between the two kinds of cobalt. However, it is more profitable to include here the curves of $\int H\,dI - Q$ against I; these are shown in figures 11 *a* and 11 *b*, in which (to avoid confusion with preceding figures) the arrows on the curves indicate the direction in which the magnetization was changed. The most prominent feature with annealed cobalt is the magnitude of the changes in $\int H\,dI - Q$, this quantity being denoted by E_T in view of the theoretical discussion given later, as I is changed from zero to I_{max} along any chosen half-cycle, compared with the change in E_T between the same limits on the virgin curve; the former is much smaller. The opposite holds for annealed nickel, where the central portions of the E_T, I curves fall below the origin.

Following the treatment given in Papers I and II, we write

$$H\,dI = \left(\frac{\partial E}{\partial I}\right)_T dI + \left(\frac{\partial E}{\partial T}\right)_I dT$$

or

$$\left(\frac{\partial E}{\partial I}\right)_T dI = H\,dI - dQ,$$

where dQ is the change in energy which appears as heat and is measured directly in our experiments, while $(\partial E/\partial I)_T\,dI$ represents the change in internal energy which is non-thermal in character. The second equation holds independently of whether the change is thermodynamically reversible or not. Hence the importance of the $\int H\,dI - Q, I$ curve lies in the fact that the slope at any point on it gives the value of (dE_T/dI), i.e. $(\partial E/\partial I)_T$ for the corresponding point on the I, H curve.

Now, the magnetization of a substance, starting from the virgin unmagnetized state, proceeds mainly by virtue of magnetically reversible 90-degree displacement followed by irreversible 180-degree displacements located chiefly on the steeper parts of the hysteresis curve. On the application of a sufficiently strong field, magnetically reversible rotations of the domain vectors set in on the upper portions of the curve beyond the "knee", and these in turn are followed by the changes in the intrinsic intensity of magnetization of the domains themselves as the applied field becomes so great that the effects of hysteresis fade into insignificance and the well-known magnetocaloric effect appears.

Starting with the material magnetized to technical saturation, a reduction in the field at first permits the magnetically reversible vector rotations to take place, followed at lower fields by 90-degree boundary displacements which are mainly reversible. In the region $I=I_r$ the curve becomes steeper and, in the main, irreversible 180-degree boundary displacements predominate. Beyond $H=+H_c$ the substance is remagnetized in the opposite sense, and the more important changes are essentially the same as for the steeper outer portions of the virgin curve. Of course, the above division of processes is approximate only, since reversible and irreversible changes must occur simultaneously in all parts of the cycle.

Bates and Weston found that $(\partial E/\partial I)_T$ was zero when $I=I_r$, but this result is definitely not applicable to the case of unannealed cobalt, and only when the metal has been exposed to a high field is there indication that it holds in the case of annealed cobalt. They also found that, in general $(\partial E/\partial I)_T$, was negative in the closed cycle whenever the main magnetic processes involved were reversible, and positive when these were irreversible. The fields used in the present work were manifestly too low to produce marked reversible rotations of the domain vectors, so we find that there is no change in the sign of $(\partial E/\partial I)_T$ when I is near its maximum value.

Cobalt is peculiar in that the cast material, which, presumably, has been quenched, is frequently in the form of face-centred cubic crystals, while cobalt annealed *in vacuo* and cooled slowly possesses hexagonal crystal structure, although a purely hexagonal specimen is most unlikely. Consequently, the magnetostriction properties of these two kinds of material are markedly different. Thus cast material, according to Nagaoka and Honda (1902) behaves in the reverse manner to iron, i.e. in weak longitudinal fields the material shortens, but, as its magnetization is increased, i.e. in fields of about 800 oersteds and upwards, it lengthens. On the other hand, annealed cobalt behaves very much like nickel, i.e. it shortens on magnetization in longitudinal fields of all intensities. The graph of dl/l against H for the two materials intersect when H is about 400 oersteds.

We might therefore expect the virgin Q, H curves for the two materials to be markedly dissimilar, particularly in the region of low fields. In practice, however, very little thermal change at all occurs until fields of over 50 oersteds are reached, and both for annealed and unannealed cobalt there is practically a linear fall of temperature between 100 and 400 oersteds, the rate of fall for the unannealed being three times as great as for the annealed. Unfortunately, as the magnetostriction of cobalt is not isotropic, we cannot base reliable deductions on the differences in the virgin curves with our present meagre knowledge.

One important extension of the work remains to be made. Cobalt forms interesting alloys with copper containing up to 8 per cent of the latter metal. These alloys have a face-centred cube structure. It is proposed to obtain some of these alloys and to make measurements with them on the above lines, in order to try to find how much the differences between the behaviour of cobalt, on one hand, and iron and nickel, on the other, may be attributed to differences in crystal structure.

§ 5. ACKNOWLEDGMENTS

Much of the apparatus used in this research was purchased from a grant made to L.F.B. by the Government Grant Committee of the Royal Society. Our thanks are due to Mr. L. H. Goris of Messrs. Brandhurst & Co. Ltd. for the supply of cobalt. We record our warm thanks to Mr. E. G. Harrison for allowing us to quote data obtained by him.

REFERENCES

Bates, L. F. and Weston, J. C., 1941. *Proc. Phys. Soc.*, **53**, 5.
Bates, L. F. and Healey, D. R., 1943. *Proc. Phys. Soc.*, **55**, 189.
Kaya, S., 1928. *Sci. Rep. Tôhoku Univ.*, **17**, 1157.
Nagaoka, B. H. and Honda, K., 1902. *Phil. Mag.*, **4**, 45.
Okamura, T., 1936. *Sci. Rep. Tôhoku Univ.*, **24**, 744.
Weston, J. C., 1940. Thesis (Ph.D. London).

Note added 12 *November* 1946

In view of the way in which crystal structure and magnetostriction must play a part in determining the variation of E_T with I, it is essential to get as much information as possible about these factors. We have been very fortunate in obtaining an X-ray examination of the two specimens of annealed and unannealed cobalt which was kindly made for us by Dr. H. Lipson of the College of Technology, Manchester. He reported that he took photographs in an ordinary single-crystal camera, the specimens being offset so that the rays hit only one edge. For this reason the specimen could not be completely rotated and was oscillated through an angle of 15 degrees.

Dr. Lipson found that the hard-drawn specimen was mainly hexagonal, but there was a slight amount of cubic phase present. Some of the lines due to the hexagonal form were broadened somewhat, as found by Dr. Lipson and Miss Edwards in their work on cobalt (*J. Inst. Metals,* **69**, 177 (1943)). There was evidence of slight preferred orientation in the (0002) reflexion, but this was not investigated in detail. The whole pattern was rather diffuse, owing, presumably, to residual stresses.

The annealed specimen showed rather better line definition and there was no evidence of preferred orientation. There was possibly a little more cubic phase present, but it was not very definitely indicated.

We therefore conclude that the specimens are nothing like as diverse in crystal structure as one might anticipate from their history or from the results reported in our communication.

INVESTIGATIONS ON ABSORPTION HYGROMETERS AT LOW TEMPERATURES

By E. GLUECKAUF,
Durham Colleges in the University of Durham

MS. received 30 *October* 1946

ABSTRACT. A method has been developed for the calibration of hygrometers and for measuring their response at low temperatures. Investigations have been made on gold-beater's skin as a hygrometric element, and on the electrolytic hygrometer described by Dunmore. Improvements concerning the response of the latter are suggested. A new optical hygrometer, based on the interference of light reflected by thin hygroscopic films, is described in detail ; its advantages are temperature-independent calibration and satisfactory response down to −60° c.

§ 1. INTRODUCTION

UNTIL about 1939, humidity measurements at high altitude were predominantly based on the hair hygrometer, which in this country was employed in the form of the Dines Meteorograph. A detailed study of the behaviour of hair at low temperatures (Glückauf, 1944) permitted the interpretation of high-altitude soundings and the unsuspectedly low humidities in the stratosphere calculated in this way (Glückauf, 1945 a) have since been confirmed by the much superior direct method of Dobson, Brewer and Cwilong (1946).

Since 1939 a number of new hygrometric elements have come into use, in particular the electrolytic hygrometer by Dunmore, used in U.S.A. (Dunmore, 1939; Diamond *et al.*, 1940), and the gold-beater's skin hygrometer which was being used in this country for radio-sondes. It was therefore desirable to investigate in some detail the working conditions of these absorption hygrometers at low temperatures, as had been done in the case of hair.

To do so, it was first necessary to develop a technique of controlling relative humidity below −15° c., the lack of which was generally admitted to be a handicap in the development and calibration of hygrometric instruments (see Diamond *et al.*, 1940, pp. 358 and 362).

§ 2. APPARATUS

The principal features of the testing apparatus are shown in figure 1.

Air from a compressor, after passing over solid caustic soda, was dried by freezing out in a copper coil immersed in a bath of trichlorethylene cooled with solid CO_2. Any remaining moisture was removed by further drying with silica gel at a pressure of 130 atm. in the steel cylinder A. (More economical methods for the production of completely dry air could have been devised, e.g. passing the compressed air through a liquefier with inefficient heat exchanger, so that the non-liquefied dried air escapes at a temperature of about −60° c. This would have obviated all the other apparatus and at the same time would have supplied completely dry air at a low temperature. But such apparatus could not be built under war-time conditions.) After passing through a reducing valve, the completely dried air is divided into two air currents of known mass ratio by means of the distribution stopcock B. The two air currents are cooled down in two coaxial coils immersed in a low-temperature bath, and one of them is saturated (with respect to ice) at the bath temperature. They are then reunited in the testing chamber and give an air current of well defined relative humidity which is identical with the percentage of air passed through the saturation coil.

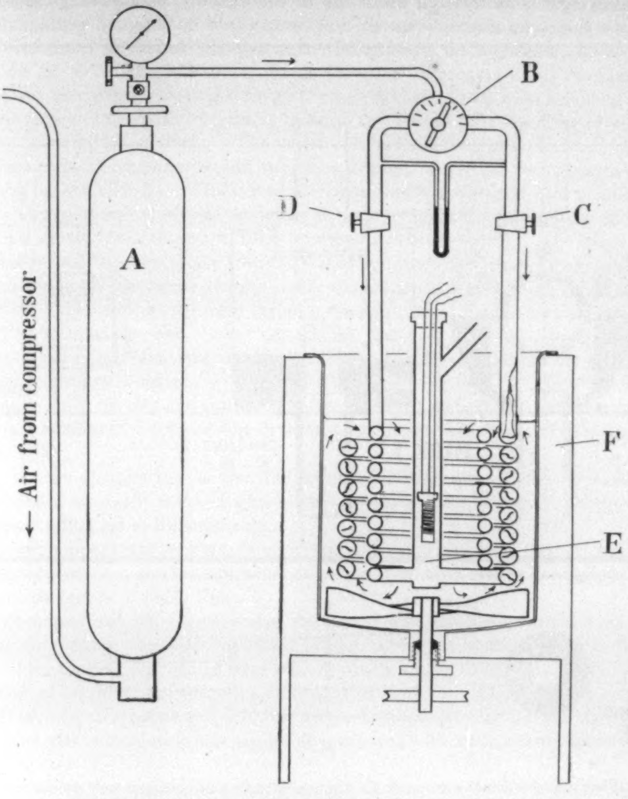


Figure 1. Apparatus for testing hygrometers at low temperatures.

A. Drying cylinder filled with silica gel.
B. Distribution stopcock.
C.
D. } Stopcocks for pressure equalization.
E. Testing chamber for hygrometric elements.
F. Lagging.

The distribution stopcock is shown in figure 2. The conical part contains nine equal holes of about 1 mm. diameter, placed at 22·5° to each other, and leading into the centre of the stopcock. If in position, one of the holes was always covered, leaving eight holes through which the air could pass. By adjusting the position of the stopcock, nine different air-flow ratios could be obtained without altering the total amount of air passing. In this way the following nominal relative humidities (with respect to ice) were obtained: 0, 12·5, 25, 37·5, 50, 62·5, 75, 87·5, 100%, which cover a range sufficient for calibrating purposes. (Due to constructional inaccuracies, the actual flow ratios, as determined by separate flow measurements in both lines, differed somewhat from the nominal ones.)

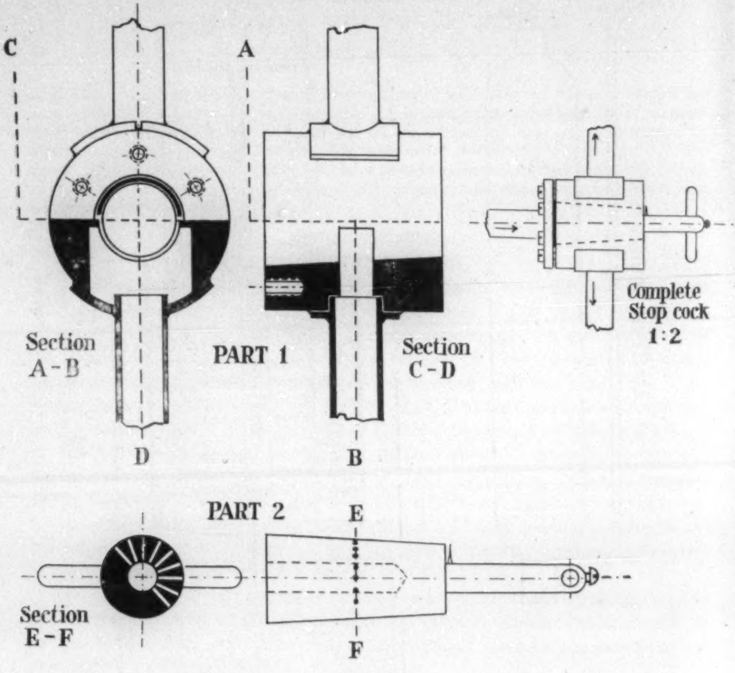


Figure 2. Distribution stopcock for dividing air stream according to well defined ratio. (1) Barrel, (2) key. Complete stopcock in reduced size.

In some earlier experiments a different arrangement using several glass stopcocks and calibrated capillaries was used which was based on the same principle but gave a greater number of combinations. It could not, however, be used for large quantities of air on account of the much higher pressures involved.

To guarantee the correct working of the distribution stopcock it was necessary to ensure equal pressure on the two outlet sides. Friction in the coils of the cooling tank caused considerable pressure differences, if the quantities of air passing through the two coils were very different (e.g. in the cases of 12·5 and 87·5% ralative humidity). To avoid errors of this kind additional flow resistance could be provided by two valves C and D, one of which was adjusted until the mercury manometer showed no difference of pressure. The arrangement made it possible to produce large well defined changes of humidity in the testing chamber within seconds.

The two air currents passing through the valves C and D were cooled down to the temperature of the bath by passing through two helical coils of tubing of $\frac{5}{8}$ in. and $\frac{1}{2}$ in. diameter respectively. The wider coil had a ribbon of cotton bandage held at the ends by copper wire and stretched through nearly the entire length of the tubing (about 15 feet). The cotton was occasionally moistened with distilled water; this arrangement proved sufficient to saturate the air completely, even at the highest speeds, with respect to ice at temperatures below $-30°$. The method described proved superior to the freezing out of surplus water vapour from moist air, as it avoided the formation of snow inside the tube, which at the high air speeds used would have been driven right through the apparatus. In the case of higher temperatures, however, the "wet" air stream was first moistened by passing through a wash-bottle filled with a suitable solution, so as to prevent the wet ribbon losing its water too quickly.

The two coils, for dry and saturated air respectively, were connected eccentrically to the bottom of a tube of 1 inch internal diameter, which served as testing chamber (E).

The eccentric connection is essential, as otherwise the two air streams do not mix efficiently, especially at very high and very low humidities, where the velocities in the two feeding tubes differ greatly.

The low-temperature bath was filled with trichlorethylene. As the bath had a fairly large capacity, a good temperature constancy was obtained by continuously adding small pieces of solid CO_2.

To obtain a uniform temperature, the liquid was vigorously agitated by a motor-driven stirrer, the shaft of which passed through a stuffing gland in the bottom of the tank. The blades were bent in such a way as to produce a very fast circulation of liquid in the direction indicated in figure 1. The bath container was held in a wooden frame and was only moderately insulated (F), as the transfer of heat through the insulation was negligible compared with the heat introduced by the passing air.

The units to be tested were either inserted directly into the testing chamber (as shown in figure 1, with an electrolytic hygrometer) or, if very high air velocities were required, they were held in ebonite holders which narrowed down the cross-section of the air flow. In the experiments with air velocities of about 100 miles/hour (45 m./second) the cross-section of air flow was thus narrowed down to 1 sq. cm.

§ 3. THE GOLDBEATER'S-SKIN HYGROMETER

(i) *Calibration curve of the G.B.S. at low temperatures*

Calibration curves have been taken for relative humidities from 0 to 100% at temperatures ranging from $+18°$ c. to $-65°$ c. These show that, at temperatures below zero, the "G.B.S." indicates relative humidity with respect to supercooled water, similar to the hair hygrometer. Table 1 gives the readings of the G.B.S. hygrometer at −21° c. in terms of the calibration at room temperature.

Table 1

% R.H. of air used relative to ice	% R.H. indicated according to room-temperature calibration	% R.H. of air used, relative to supercooled water
100	81, 78·5, 79	81·0
85	70	69·8
67·5	54, 55	54·6
57·5	46	46·6
52·5	41·5	42·5
42·5	34, 33	34·4
32·5	26, 27·5, 27	26·3
15	11·5, 12·5	12·1

In another series of experiments the relative humidity indicated by the G.B.S. was measured in air fully saturated over ice. The figures are given in table 2. Both tables show clearly that calibrations at room temperature can be applied without corrections to all temperatures, provided that the relative humidity so obtained is related to supercooled water and not to ice. The term relative humidity", or "R.H.", is therefore used in future only with respect to supercooled water, unless the reference to ice is specially stated.

Table 2

Temperature (°c.)	% R.H. indicated according to room-temperature calibration	% R.H. of air used, relative to water
+ 0·8	100	100
− 2·3	97·5	97·9
− 6·5	93·5	94·0
−10·5	90·5	90·5
−15·5	85·1	86·1
−19·2	82·6	82·9
−22·0	80·4	80·7
−25·6	77·5	77·7
−30·0	73·7	74·2
−33·5	72·5	71·6
−36·5	70·0	69·6
−39·5	67·1	67·5
−42·5	65·7	65·6
−65	53	52·6

(ii) *Hysteresis effects after exposure to low humidities*

Like the hair hygrometer, the G.B.S. hygrometer also suffers from a hysteresis effect after exposure to relative humidities lower than 30% with respect to water (see figure 3). This hysteresis effect takes place also at low temperatures; curves similar to those of figure 3, some of them almost identical, have been observed at −21° c. and −29° c. The effect of exposure to low humidities disappears only at humidities higher than 70% R.H. with respect to water, and this also occurs at −29° c. As, below −40° c., ice saturation corresponds to less than 70% R.H., the hysteresis effects are no longer relieved by occasional high relative humidities, and considerable uncertainty as to the interpretation of the readings must be accepted at these low temperatures.

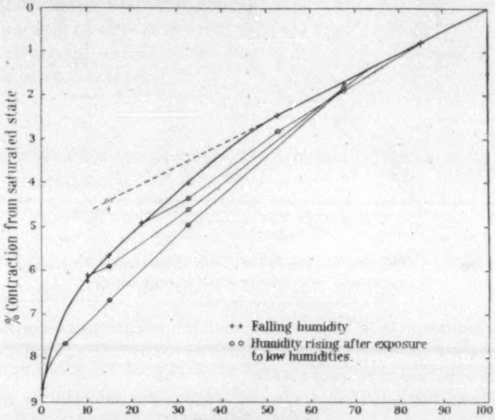


Figure 3. Calibration curve of gold-beater's skin at 12° c. (showing hysteresis curves).

Exposure at −40° to relative humidities as high as 35% R.H. (ice) may cause hysteresis errors in subsequent readings as high as 10% R.H. (ice), and this error may be doubled by exposure to lower humidities at lower temperatures.

(iii) *Response of gold-beater's skin at atmospheric pressure*

A large part of this section deals with the lag of the G.B.S. In describing this response, use is often made of the conception of the half-change time τ, which is the time required to complete the first half of the change in indicated humidity after a sudden change of the relative humidity of the surrounding air. The reason for using the half-change time is that the lag curve cannot be represented by a simple exponential function like many other reactions. Under these circumstances the "half-change time" gives a much clearer perception of the response than the more precise mathematical expressions using a "response constant" k of a well defined function. The use of k will therefore be restricted to the parts dealing with the theoretical interpretation of the lag curve.

The response has been measured for changes at medium humidities at temperatures between + 18 and − 68° c. and at air velocities ranging from 1·2 to 34 metres per second. In figure 4 the logarithms of the half-change times are plotted against the reciprocals of absolute temperatures (T), resulting in parallel lines for the various air velocities. It is to be noted that the temperature coefficient of the response constant k (inversely proportional to τ) is only slightly smaller than that of the saturation pressure of liquid water.

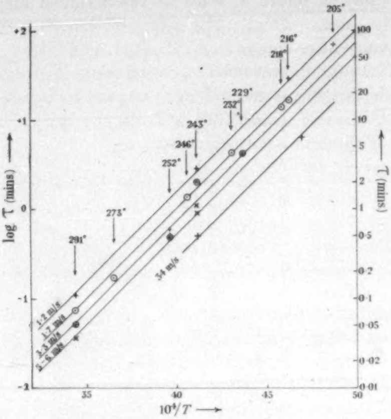


Figure 4. Response of gold-beater's skin at low temperatures and various air velocities (τ=half-change time).

Table 3 shows the half-change time at − 30° c. for different air velocities (u).

Table 3

u (m./sec.)	τ (min.)	$\frac{-d \log \tau}{d \log u}$
1·2	2·5	
		1·2
1·7	16·5	
		0·42
3·3	1·25	
		0·38
6	1·00	
		0·40
34	0·5	

It can be seen that, apart from the lowest value of u, τ is inversely proportional to $u^{0·4}$, which may be taken as an indication that the lag is almost entirely due to the diffusion of water molecules through the boundary layer surrounding the G.B.S. This is also borne out by the quantitative agreement between the half-change times found experimentally and those theoretically calculated on the basis that the lag is due to the boundary layer.

§4. CALCULATION OF THE LAG CURVE

Let dA/dh be the change of weight with relative humidity (h) per cm^2 of film or skin ($h=1$ at water saturation), so that $dA/dh=a\,.\,A_0$, where A_0 is the weight of film in gr./cm^2 and a may have different values in different regions of relative humidity.

$D=\dfrac{250}{P}\cdot\left(\dfrac{T}{273}\right)^{1\cdot 75}$ = diffusion constant of water vapour in air at a pressure of P millibars.

C = saturation concentration over water in gm./c.c.

δ = thickness of boundary layer.

n = number of sides of the film exposed to air (1 or 2).

As the change of weight of the film is due to the diffusion of water vapour through the boundary layer,

$$\frac{dA}{dt}=\frac{n\,.\,D\,.\,C}{\delta}(h_1-h)$$

at the time t after a sudden alteration of the humidity from h_0 to h_1. The change of indicated humidity (h) will then be

$$\frac{dh}{dt}=\frac{n\,.\,D\,.\,C}{\delta\,.\,a\,.\,A_0}(h_1-h)$$

or, integrated for changes of humidity not large enough to affect the value of a,

$$\frac{h_1-h}{h_1-h_0}=\exp.-\frac{n\,.\,D\,.\,C}{\delta\,.\,a\,.\,A_0}\,.\,t. \qquad \ldots\ldots(1)$$

Unfortunately, the simplicity of the function (1) is affected by δ not having a constant value along the length of the film, if the latter is in a position parallel to the air flow. The only case permitting of an exact evaluation is the case of a laminar air flow along the exposed film, when

$$\delta=4\cdot 5\sqrt{\frac{\nu\,.\,x}{u}}\ \text{cm.},$$

where ν = kinematic viscosity, x = distance upwind from edge of film in cm., and u = air velocity in cm./sec.

Under these conditions the indicated change of relative humidity (h) is given by

$$\frac{h_1-h}{h_1-h_0}=\frac{1}{L}\int_0^L e^{-kt/\sqrt{x}}\,.\,dx, \quad \text{where} \quad k=\frac{n\,.\,D\,.\,C\,.\,\sqrt{u}}{4\cdot 5\,.\,a\,.\,A_0\,.\,\sqrt{\nu}}, \qquad \ldots\ldots(2)$$

and where L is a length of the film or skin under consideration. The value of the integral (2) can be evaluated by means of Soldner's function (see Edwards, *Integral Calculus*, p. 334)

$$\frac{1}{L}\int_0^L e^{-kt/\sqrt{x}}\,.\,dx=\left(\frac{kt}{\sqrt{L}}\right)^2.\,\mathrm{li}\,(e^{-kt/\sqrt{L}})-\left(\frac{kt}{\sqrt{L}}-1\right).\,e^{-kt/\sqrt{L}}$$

in which li is the logarithmic integral. It is shown in curve B of figure 5 as a function of $kt/\sqrt{L}$, so that for a given value of $k/\sqrt{L}$ the time required for a certain change, e.g. 50% or 90%, can be directly obtained from this curve. It can be seen that curve B differs considerably from the functions exp. $-2kt/\sqrt{L}$ (curve A) and that it gives expression to the fact that the forward parts of the film or skin reach equilibrium earlier than those further down stream.

In applying equation (2) to the experimental data of the G.B.S. hygrometer used for the tests, the following numerical values must be employed:

$$n=2, \qquad L=3\text{ cm.}, \qquad A_0=1{\cdot}35\ .\ 10^{-3}\text{ gm./cm}^2,$$
$$a=0{\cdot}2 \text{ (near } h=50\% \text{ R.H.)}.$$

The other physical data are given in table 4.

Table 4

T (°c.)	$10^6.C$ (gm./c.c.)	D	ν	u (cm./sec.)	$\frac{10^3k}{\sqrt{L}}$ (sec.$^{-1}$)	τ calc. (min.)	τ obs. (min.)	$\frac{\tau \text{ calc.}}{\tau \text{ obs.}}$
273	4·84	0·25	0·19	200	37	0·18	$\sim$ 0·2	$\sim$1
230	0·135	0·18	0·10	170	1·0	6·6	5	1·3
				330	1·4	5·0	3·6	1·4
				3400	4·3	1·5	1·3	1·2
216	0·031	0·16	0·09	120	0·18	37	29	1·3
214	0·026			3400	0·79	8·5	6	1·4

For the purpose of comparing the calculated lag values with those found experimentally and plotted in figure 4, it will again be an advantage to compare the times of the half-change τ. Not only can this point be determined experimentally with greater accuracy than points nearer the completed equilibrium, but, as will be seen shortly, there are difficulties in reducing the lag curves obtained experimentally to fundamental constants like $k/\sqrt{L}$. As shown in table 4, the calculated τ is larger than the observed one by a factor of 1·3 on the average, but the two τs are not strictly comparable.

On account of the limited amount of air at our disposal (less than 10 c. ft./min.), the experimental investigation was carried out in a very narrow channel of X-section 0·4 × 2·5 cm², and the flow was thus not laminar, but highly turbulent at the entrance of the channel.

Tables 5 *a* and 5 *b* show typical lag curves at a temperature of $-29°$ c. taken at an air velocity of 4·2 m./sec. The humidity range chosen (from 90% to 57% R.H. with respect to ice) corresponds to medium humidities with respect to water, where the value of *a* is very constant. Comparison of the third columns in both tables show that the G.B.S.—unlike the hair hygrometer—follows the same response function independent of the direction of the humidity change.

It can be seen from tables 5 *a* and 5 *b* that the lag curves obtained experimentally are neither simple exponential like curve A, nor do they follow curve B, which is based on the assumption of laminar flow (equation (2)). They correspond to a curve of type C (figure 5). Curve C is based on the assumption that the thickness δ of the boundary layer is proportional not to $\sqrt{x}$ but to x, resulting in the equation

$$\frac{h_1-h}{h_1-h_0}=\frac{1}{L}\int_0^{x=L} e^{-\frac{kt}{bx}}\,.\,dx, \quad \text{where } k=\frac{n\,.\,D\,.\,C\,.\,\sqrt{u}}{4{\cdot}5\,.\,a\,.\,A_0\,.\,\sqrt{\nu}}, \quad \ldots\ldots(3)$$

and where the constant b may have values up to $1/\sqrt{L}$, which means that δ cannot exceed the value it would have under laminar flow conditions. Curve C has been constructed for $b = 1/\sqrt{L}$. Equation (3) can be integrated by using again Soldner's "logarithmic integral",

$$\frac{1}{L}\int_0^L e^{-\frac{kt}{bx}} \,.\, dx = e^{-\frac{kt}{bL}} + \frac{kt}{bL} \,.\, \mathrm{li}\left(e^{-\frac{kt}{bL}}\right).$$

Table 5 *a*. $T = -29{\cdot}0°$ c.

t (min.)	h observed	$\frac{h_1 - h}{h_1 - h_0}$	$k/\sqrt{L}$ according to "A"	$k/\sqrt{L}$ according to "B"	$k/\sqrt{L}$ according to "C"
0	90% = h_0	1·000			
0·25	83·4	0·798	0·48	0·48	0·28
0·5	78·9	0·660	0·41	0·45	0·29
0·75	75·5	0·557	0·39	0·44	0·30
1	72·8	0·472	0·38	0·44	0·30
1·5	69·0	0·356	0·35	0·43	0·305
2	66·3	0·274	0·33	0·41	0·30
3	63·1	0·175	0·29	0·38	0·29
4	61·5	0·125	0·26	0·34	0·29
5	60·4	0·090	0·24	0·32	0·27
6	59·0	0·047	0·25	0·34	0·29
10	57·5% = h_1				

Table 5 *b*. $T = -28{\cdot}5°$ c.

t (min.)	h observed	$\frac{h_1 - h}{h_1 - h_0}$	$k/\sqrt{L}$ according to "A"	$k/\sqrt{L}$ according to "B"	$k/\sqrt{L}$ according to "C"
0	57% = h_0	1·000			
0·25	65·4	0·770	0·54	0·58	0·32
0·5	69·2	0·614	0·46	0·54	0·36
0·75	71·5	0·555	0·39	0·45	0·30
1	74·2	0·475	0·37	0·44	0·30
1·5	78·4	0·349	0·36	0·43	0·31
2·25	82·2	0·232	0·33	0·41	0·31
3	84·6	0·163	0·30	0·40	0·31
4	86·2	0·109	0·28	0·37	0·32
5	87·6	0·070	0·27	0·36	0·30
9	90% = h_1				

As the actual lag curves agree in form with the curve from equation (3) (see column 6, tables 5 *a* and *b*), it must be concluded that, under the experimental conditions, the boundary layer along the G.B.S. is affected by the initial turbulence and has a more or less linear form. This would also explain the results of table 4, which show a smaller value for the experimentally found half-change time than that obtained under the assumption of laminar flow.

§ 5. THE RESPONSE OF GOLD-BEATER'S SKIN AT DIFFERENT RELATIVE HUMIDITIES

Figure 6 shows the response constant $k/\sqrt{L}$ ($=0{\cdot}28/\tau$) for small changes at various humidities at 18° c. The curve exhibits a marked maximum between 50 and 65% R.H. This behaviour is quite different from that of the hair hygrometer, for which the half-change times decrease throughout towards higher humidities. Figure 7 shows that this phenomenon is substantially the same at lower temperatures (−21° c. and −30° c.). The maximum for −21° c. lies again in the same region, if relative humidity is taken with respect to supercooled water; at −30° c. the falling branch is unobtainable on account of ice saturation.

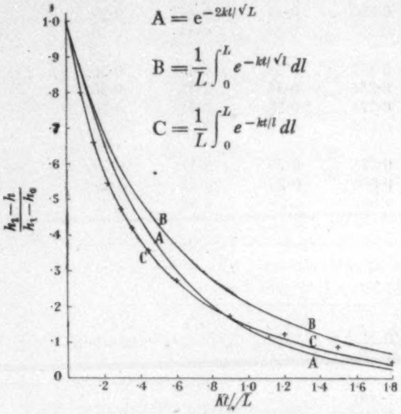


Figure 5. Calculated reponse curves of gold-beater's skin and observed response (+).

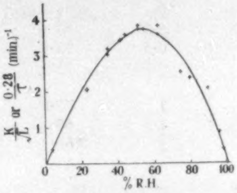


Figure 6. Response in different regions of relative humidity at 0° c.

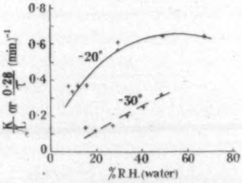


Figure 7. Response in different regions of relative humidity at −20° c. and at −30° c.

§ 6. CONCLUSIONS CONCERNING GOLD-BEATER'S SKIN

While gold-beater's skin is much faster in response than hair, it is still slow at temperatures below −40° c., as can be seen from the data of figure 4, which give the half-change times at medium relative humidities where the response is a maximum. At low humidities it is much slower still. It would even appear from figures 6 and 7 that the response ceases at relative humidities approaching zero. Observations both in the laboratory and in the stratosphere (by Mr. A. W. Brewer) confirm that gold-beater's skin loses all response after prolonged exposure to zero humidity and that it only regains it after exposure to high humidities above −15° c.

Whether a reduction of the atmospheric pressure would bring an improvement of the normal response could not be ascertained experimentally. Rough calculations under the assumption that the internal diffusion coefficient of gold-beater's skin is similar to that of hair would lead to the conclusion that, while reduction of pressure would improve the response at medium to high humidities, it would not have a great effect below 20% R.H., where a reduction of pressure would make internal diffusion the dominant factor.

The hysteresis effect after exposure to humidities below 20% R.H. is less than in the case of hair. However, it is still far from negligible and is likely to affect the accuracy of the readings.

These results show that, while gold-beater's skin is an improvement on hair, its limitations at low humidities are considerable.

§7. THE ELECTROLYTIC HYGROMETER

Tests similar to those with gold-beater's skin were made with units of the electrolytic hygrometer of Dunmore (1939), which is based on the resistance of a thin electrolyte film containing lithium chloride. These tests appeared to be all the more necessary as the instrument was apparently used at low temperatures without any calibration except at ice saturation, other humidities being extrapolated from the behaviour at room temperature.

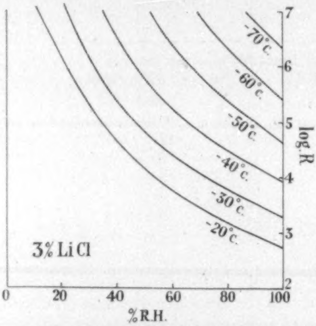


Figure 8. **Resistance (R) of Dunmore unit with 3 % LiCl at different temperatures and humidities.**

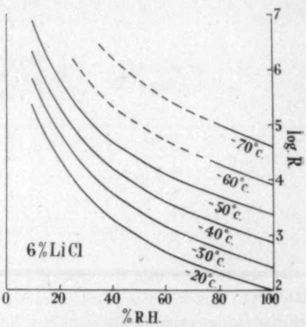


Figure 9. Resistance (R) of Dunmore unit with 6% LiCl at different temperatures and humidities.

The resistance measurements at low temperatures and at different humidities, which were carried out with units obtained from U.S.A., showed immediately that this procedure was not admissible. The temperature coefficient of the resistance varies enormously at different relative humidities, as can be seen from figure 8, for a unit made from a 3% LiCl solution. (For composition of solutions see Dunmore (1939), p. 705). For a change from −20° c. to −40° c. the resistance of these units changes

at ice saturation by a factor of 14,
at 70% R.H. (ice) by a factor of 25,
at 40% R.H. (ice) by a factor of 70.

Such variations make an extrapolation from room-temperature conditions impossible, and these differences are greater still for units containing smaller percentages of LiCl, such as are used in the composite hygrometer described by Dunmore (1939), figures 2 and 3, and Diamond, Hinman and Dunmore (1940).

On the other hand, units made from solutions with more than 6% LiCl did not show this variation with humidity of the temperature coefficient of the resistance (see figure 9). This fact, as will be discussed later, makes possible the construction of temperature-independent hygrometers of this type.

The electrolytic hygrometer does not show any hysteresis after exposure to low humidities. This great advantage over hair and gold-beater's skin is probably due to the amorphous structure of the film. This, on expansion or contraction, does not set up the internal stresses responsible for a hysteresis effect (see Barkas, 1942) which would occur in semi-crystalline structures like cellulose, hair and gold-beater's skin.

Response of the Dunmore hygrometer

The response of the Dunmore units which, on account of the form of the humidity chamber, had to be tested with the air flow parallel to their axes, is shown in table 6 for a unit of 20 wire turns made from 3% LiCl solution at a temperature of −28° c.

Table 6

Air velocity (m./sec.)	Half-change times for changes from : 57% to 90% R.H. (min.)	90% to 57% (ice) (min.)
0·4	4·5	12
2·7	2·0	5
5·2	1·1	3
6·7	0·7	-
44	0·2	0·5

It can be seen that the change from low to high humidity is about 2·5 times faster than that in the opposite direction, and that the response depends on the air velocity (u) approximately as $u^{0\cdot 65}$. The dependence on the air velocity indicates that the lag is due to the diffusion of water vapour through the boundary layer. The reason for the influence of the direction of the humidity change can also be easily understood. The various parts of the electrolytic hygrometer unit may be considered as "resistances in parallel", so that the total resistance is substantially determined by the parts of lowest resistance, i.e. of highest moisture content. During the change from low to high humidity, the total resistance is thus at first governed by the most exposed parts, while during the opposite change the resistance depends largely on the least ventilated parts of the unit, and consequently shows less response during the initial changes.

The variation of the response with temperature is given in table 7 for changes from low to high humidity at an air velocity of 44 m./sec. (unit: 20 turns, 6% LiCl).

Allowing for differences in flow velocity, temperature and arrangement, these times are in line with the figure given by Diamond, Hinman, Dunmore and Lapham, who find a half-change of 0·12 minute at 5 m./sec. at 0° c.

As these half-change times are increased by a factor of 6 for ventilation speeds of 5 m./sec. (normal for radio-sonde work) and by another factor of 2·5 for changes from high to low humidity, it is apparent that the lag of this hygrometer, when

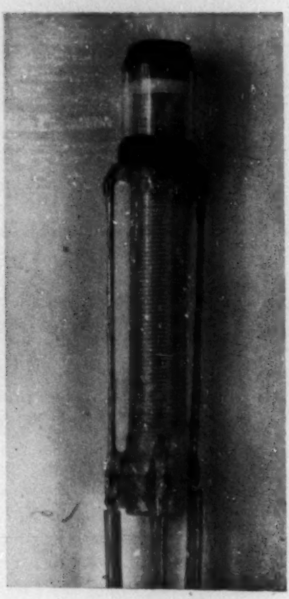


Figure 10. Electrolytic hygrometer with temperature compensation and wide range of humidities (circuit, see figure 11).

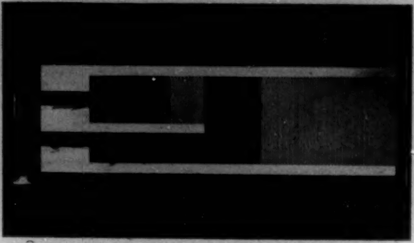


Figure 15. Electrolytic hygrometer element (for temperature compensation, made by Scientific Instrument Research Association).

used in the radio-sonde, cannot be neglected at temperatures below −40° c. It may, however, be expected that in this case considerable improvements may be caused by the low density of air in the upper atmosphere.

Table 7

Temperature (°c.)	Half-change (min.)
−20	<0·1
−30	0·2
−40	0·7
−50	2
−60	appr. 6
−70	appr. 25

§ 8. MODIFICATIONS OF THE ELECTROLYTIC HYGROMETER

Improvements were directed towards designing a temperature-independent electrolytic hygrometer (thus eliminating the necessity to calibrate every instrument at low temperatures) and towards increasing the response.

The first aim was achieved by compensating the temperature effect by means of a similar electrolytic resistance kept at ice saturation. This could be done for units made from 6% LiCl solution, where, as stated before, the temperature effect is the same at all humidities.

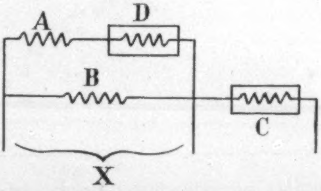


Figure 11. Circuit of electrolytic hygrometer with temperature compensation.

To render the scale readings more linear, the compensated hygrometer, shown in figure 10, was composed of four resistances, all made from polyvinylacetate with 6% LiCl, as follows:

Resistance A = 20 turns at 16 turns per inch,
B = 2 turns at 16 turns per inch,
C = 1 ring, 1/16 inch wide,
D = 1 ring, 1/8 inch wide.

These arrangements had the result that A : B : C : D = 1 : 7 : 42 : 21 if all were kept at the same humidity. In the hygrometer, the circuit of which is shown in figure 11, A and B were exposed to the air, while C and D were kept in an atmosphere saturated with respect to ice. The saturation was produced by keeping C and D in an annular closed space containing a ring of filter-paper soaked with saturated K_2SO_4 solution. This produced ice saturation below −5° c., and at the same time prevented the "flooding" of the LiCl film at room temperature which would occur with pure water.

The resistance ratio X/C was found to be independent of the temperature between $+5°$ and $-60°$ c. where X and C are the resistances of the unit and the compensator respectively. Figure 12 shows a series of calibrations, carried out over a fortnight, with the compensated hygrometer shown in figure 10. In this test the relative humidities (with respect to ice) were measured with an early model of the dew-point hygrometer by Dobson, Brewer and Cwilong, kindly lent by the Meteorological Office of the Air Ministry. It is difficult to say whether the scatter of the calibration points was due to the electrolytic hygrometer or to the difficulty in operating this dew-point instrument, which was not as sensitive as later models. But apart from this, the constancy of the resistance ratio was well established, thus providing the basis for electrolytic hygrometers with temperature-independent calibration.

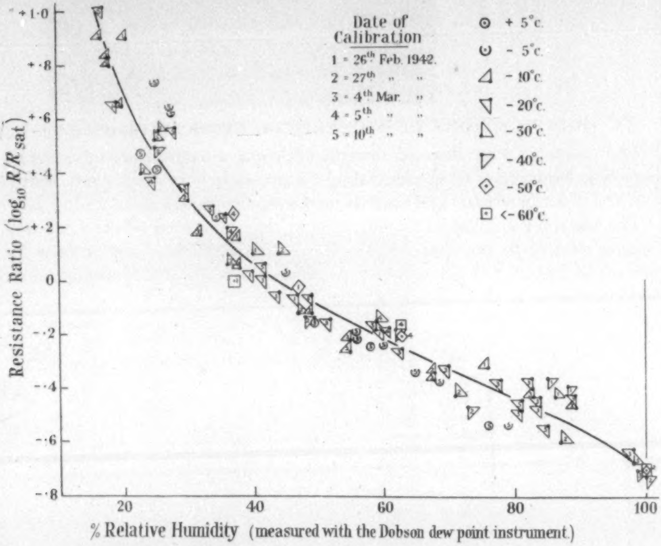


Figure 12. Calibration curve of compensated electrolytic hygrometer.

To increase the response of the electrolytic units, the electrodes were platinized directly on to the glass holders, as shown in figure 13. This unit was produced by painting the electrode pattern with platinizing solution on a glass bulb with subsequent heating. These units had the advantage that much less hygroscopic material adheres to the smooth surface than was retained by the wire-covered Dunmore units. Consequently, less moisture was required to produce equilibrium, which resulted in a faster response. Figure 14 shows the response curves for this type of hygrometer unit at a temperature of −50° c. at an air velocity of 18 m./sec. The half-change times were 0·4 and 0·7 minute, according to the direction of the humidity change, an average improvement by a factor of 10 over the performance of the wire-wound units.

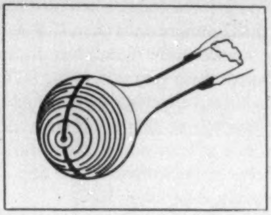


Figure 13. Electrolytic hygrometer element (platinum on glass bulb)

Still faster response was obtained with platinum-on-glass electrodes, kindly made for me by the British Scientific Instrument Research Association, which are shown in figure 15. The close spacing of the electrodes and the comparatively large electrode area made it possible to use films of less than 0·001 mm. thickness. These films gave manageable resistances, and half-change times of about 0·5 minute were obtained at −55° c. at an air velocity of 4 m./sec.

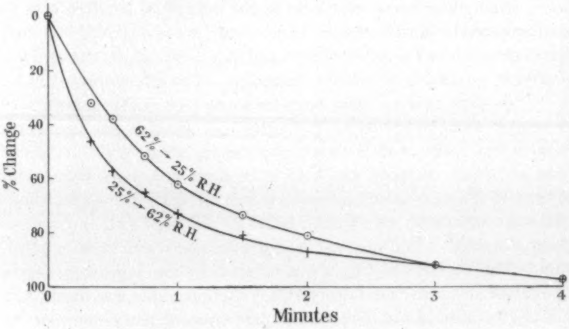


Figure 14. Response curve at −50° c. of electrolytic hygrometer unit shown in figure 13 (air ventilation 18 m./sec.).

At this point progress came to a standstill through technical difficulties the cause of which was not recognized until very much later. It was found that the platinum-on-glass units showed a slow but continuing change in calibration with time, which affected not only the absolute value, but also the temperature coefficient of the resistance. Systematic investigations revealed that the films behaved as if LiCl was constantly removed from the electrolyte.

Nor was this effect of a simple nature. In films with high LiCl content (6%), the process seemed to be accompanied by the formation of microscopic needles in the film. These formed only if the film was exposed to high humidities, but their appearance could be indefinitely postponed by keeping the units in a desiccator. No indication of their chemical composition was found, but similar needles were also obtained for mixtures of LiCl with other binders such as gelatine, gum arabic, and other types of polyvinylacetates when used on glass, while these needles could not be detected on the Dunmore units from U.S.A. Using coating solutions of less than 3% LiCl, no needles were found, but the apparent disappearance of LiCl was observable also with these very thin films. The cause of this seemed to elude all efforts at detection, until a specimen which had been kept for more than a year showed under the miscrocope that small cubic crystals had been formed. It then became apparent that at least one of the disturbing processes must have been the ion exchange between the lithium of the film against the sodium of the glass, the cubic crystals being most likely NaCl.

It is quite possible that, if this work should again be taken up, a technical solution may be found, e.g. by using quartz glass or fused alumina as bases for the platinum patterns. If this should result in stable films, the way would be open to constructing a very simple hygrometer of sufficiently fast response at all temperatures down to – 70° c. free from hysteresis effects at low humidities, which would not require low-temperature calibration, and which could easily be adapted for use both on aircraft and radio-sonde.

§ 9. AN OPTICAL HYGROMETER

During attempts to throw some light on the difficulties incurred with the very thinnest films used for the electrolytic hygrometer, it was observed that these films gave interference colours in reflected light and that these colours changed markedly under varying conditions of relative humidity. The effect seemed to be sufficiently characteristic to serve as the basis for a new type of hygrometer.

In order to test the behaviour of hygroscopic films at low temperature, an arrangement was used which is shown in principle in figure 16. A coated lead-glass disc of 10 mm. diameter was held at an angle in a metal frame within the central tube of the calibration apparatus shown in figure 1. Light from a 12-v. lamp (L) was condensed into a narrow beam by means of a microscope eye-piece (E) which was filled with a mixture of glycerine and water so as to absorb heat radiation. The "interfered" light was reflected by the hygroscopic surface (A) into a photocell (P) of the caesium type. A thermocouple was fixed to the back of the lead-glass disc (A) to observe any differences of temperature between the bath and the hygrometer unit, caused by absorption of light at the blackened back of the lead-glass disc.

The intensity of the light source was adjusted so as to give a definite photocell current, when the light was reflected by the back of the spring-loaded shutter (St) which, when released, would cover the hygroscopic surface and thus act as a standard reflector. By removing the box with the photocell it was possible to view directly the changes of colour of the reflected light.

Hygroscopic films of very uniform thickness were produced by spinning the solutions on discs of lead-glass of high refractive index. Extremely vivid colours were obtained, the variability depending on the LiCl content. By adjusting the concentration and viscosity of the solutions, and by varying the speed of the

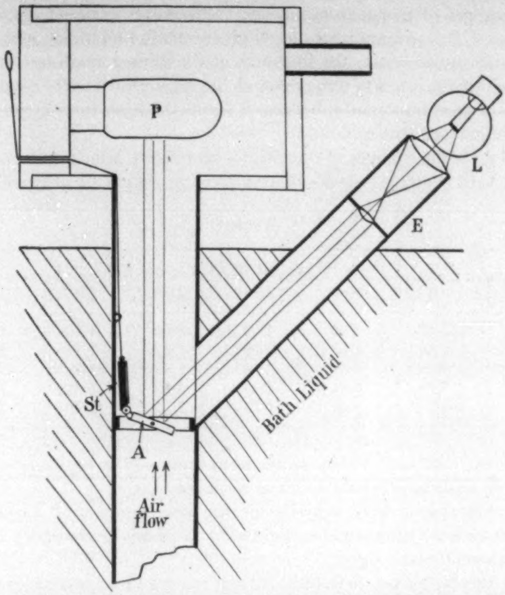


Figure 16. Arrangement used in combination with testing apparatus (see figure 1) for calibration and response measurements of "optical hygrometer" elements. L, lamp ; E, lens ; A, hygroscopic film surface ; St, standard reflector ; P, photocell.

rotating discs, any desired initial thickness of film could be produced. Examples of interference colours obtained for Solvar films on lead-glass, and their variation with relative humidity, are shown in table 8.

Table 8

% R.H.	6 % LiCl solution	1 % LiCl solution
0	Yellow	White
10	Orange	
20	Purple	
30	Blue	Yellow
40	Blue-green	
50	Yellow-green	Orange
60	Orange	
70	Red	Red
80	Blue	Purple
90	Yellow	Blue
96	Red	Green

A more quantitative investigation of the swelling properties was based on the intensity changes of monochromatic light reflected from the hygroscopic film on lead-glass. It was found that at any given relative humidity with respect to liquid or supercooled water, the thickness of the films containing LiCl remains unaltered within the limits of error between the temperatures of + 15 and − 50° c. This fact is of great importance, because it forms the basis for a hygrometer with temperature-independent calibration.

Table 9 gives the changes of thickness with relative humidity found for films of different LiCl contents (which are here given in weight % of dry film). The thickness at 0% relative humidity is defined as unity. The region between 0 and 35% R.H. shows an approximately linear response.

Table 9

% R.H.	Weight % LiCl in dry film					
	0	8·4	15·5	21·6	27·0	35·5
0	1·00	1·00	1·00	1·00	1·00	1·00
35	1·01	1·025	1·05	1·12	1·16	1·24
52	1·02	1·06	1·10	1·19	1·28	1·32
76	1·06	1·15	1·24	1·43	1·52	1·55
96	1·17	1·33	1·78	2·15	2·30	2·56

These observations were utilized for the development of a temperature-independent optical hygrometer (see figure 17) based on the change of intensity of reflected monochromatic light.

Parallel light from a 6-v. 6-w. bulb (L) was reflected by a coated lead-glass disc (A) of ½ inch diameter to a vacuum photocell of the caesium type. Caesium photocells are particularly sensitive to red light, so that the photocell even without a colour filter indicated essentially the intensity change of red light. The monochromatism was further improved by a filter of methyl violet in front of the photocell. Light from the same source fell through an adjustable slit (S) on a second photocell, which served to compensate against changes in the light intensity of the lamp. Both photocell currents were suitably amplified, using a cathode follower arrangement (Sowerby, 1944). A method (Glückauf, 1945) for measuring directly the ratio of the illuminations of the two photocells was not used because vacuum photocells proved to be more reliable than the otherwise needed gas-filled cells. Changes due to relative humidity were read with the millivoltmeter (M) of high resistance. (For satisfactory working, resistance in the voltmeter circuit should be made a good deal higher than the cathode follower resistance.) The width of the slit S is adjusted so that at zero humidity no current flows through the voltmeter.

The change of light intensity I of monochromatic light with the film thickness F(given in terms of the wave-length) should give a cosine wave, but in practice, with incompletely monochromatic light, the amplitudes of the cosine waves gradually fall off (see figure 18). In order to compensate for the very much higher expansion of the films per unit change of relative humidity at high humidities (see table 9) it is of advantage to choose the thickness and composition of the film in such a way that the zero-humidity thickness has a value where dI/dF is a maximum (e.g. $\lambda/8$, $3\lambda/8$, $5\lambda/8$, etc.) and that the 96% saturation thickness has values with minimum dI/dF (e.g. $\lambda/4$, $2\lambda/4$, $3\lambda/4$, etc.). Two useful ranges are indicated in figure 18: from $\lambda/8$ to $\lambda/4$ and from $3\lambda/8$ to $\lambda/2$. These require materials of a composition giving, between 0 and 96% R.H., an expansion of 100% and 33% respectively.

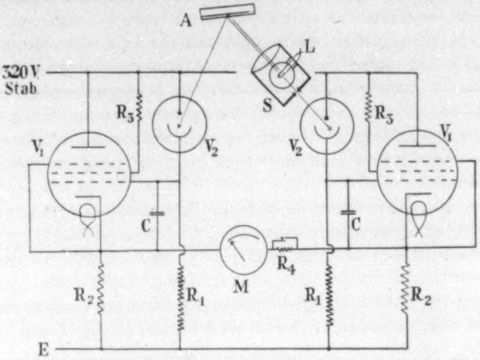


Figure 17. Arrangement and electrical circuit of optical hygrometer. A, hygroscopic surface; L, 6-watt lamp; S, shutter, adjustable. Electrical circuit: V_1, H.F. pentode; V_2, Baird vacuum photocell (Cs-type); $R_1 = 100$ M; $R_2 = 100{,}000$; $R_3 = 30{,}000$; R_4 according to resistance of meter M; $C = 0{\cdot}01$ μF.

According to table 9 this requires film materials containing 19% and 8·4% LiCl respectively in the dry film. As both films require, independently of their zero-humidity thickness, the same amount of moisture and, consequently, have the same response characteristics, there is no advantage in choosing the thinner film. As the hardness of the films increases with decreasing LiCl content, it is actually better to use the thicker film. A zero thickness of $3\lambda/8$ was therefore used throughout.

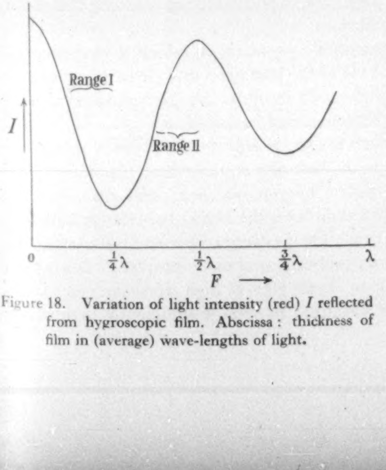


Figure 18. Variation of light intensity (red) I reflected from hygroscopic film. Abscissa: thickness of film in (average) wave-lengths of light.

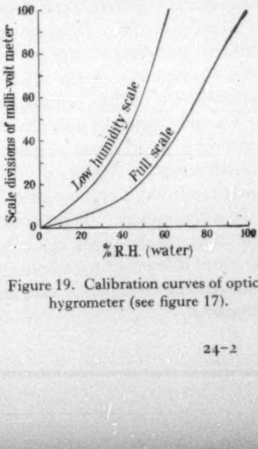


Figure 19. Calibration curves of optical hygrometer (see figure 17).

By using a range from $5\lambda/8$ to $3\lambda/4$ it would have been possible to dispense with LiCl altogether and to use pure Solvar films. But not only were the amplitudes of the I–F curve much smaller in this region, which would have reduced the proportional change in light intensity, but there were also other difficulties. It was found that the Solvar films without any LiCl showed a slight variation of their calibration curve with temperature which, in view of the wide temperature range for which the instrument was intended, could not be neglected. (Physico-chemically this phenomenon would mean that the heat of swelling of the pure Solvar films is not negligible.) For a more moderate range of temperatures, down to about −30° c., this temperature effect is almost negligible. For this purpose, the use of pure Solvar films offers considerable advantages as regards stability, for films containing LiCl are destroyed by contact with liquid water.

In order to give greater scale uniformity in different regions of humidity, two different voltmeter sensitivities were used. This is particularly useful for work at very low temperatures, where the high relative humidities with respect to water do not occur on account of ice saturation.

An instrument built on these principles with the circuit of figure 17 gave calibration curves as shown in figure 19. It can be seen that the arrangement by which the strongly expanding high-humidity part coincides with the least sensitive part of the interference curve does indeed lead to a fairly linear curve in the high-humidity region.

It was at first feared that ambiguities in the indication might arise for air humidities higher than 96% relative humidity, when the light intensity, after having reached the maximum, does again decrease. However, in practice this was not found to be the case. The radiation from the lamp heats the reflecting film to a temperature slightly above that of the surrounding air, so that under constant conditions of ventilation the relative humidity at the hygroscopic surface is always a small and constant fraction below the actual air humidity, and relative humidities above 96% do not occur at the hygroscopic surface.

At room temperature and with moderate ventilation the response of the hygrometric film is instantaneous. This is obvious, considering that the amount of moisture required for a change, e.g. from 52 to 76% T.H., is only 3 μ gr./cm², and from 35 to 52% R.H. is only 1 μ gr./cm², so that exchange with less than 1 cm. of air layer is required to reach equilibrium.

Experiments at −50° c. at atmospheric pressure, in which a ventilation of 10 m./sec. was directed against the back of the lead-glass disc (see arrangement in figure 16), gave half-change times of about 20 seconds. At high altitudes and low air densities a further increase in response may be expected.

The only technical difficulty which has so far prevented the use of the optical hygrometer for practical purposes is the lack of a suitable base material of high refractive index. When covered with a hygroscopic film, lead-glass becomes slowly tarnished, a process which proceeds faster the higher the relative humidity of the surroundings. This has the result that the calibration curve is subjected to a slow continual shift, making necessary calibrations at daily intervals. During the war years it was not found possible to obtain glass of high refractive index made from materials other than lead, nor was fused alumina of sufficiently high purity obtainable.

In the meantime, the development of a low-temperature dew-point hygrometer by Dobson, Brewer and Cwilong, with its obvious advantage of not requiring calibration, has somewhat reduced the interest in absorption hygrometers for high-altitude hygrometry. In particular, the very low relative humidities which are found above the tropopause (Glückauf, 1945 a; Dobson *et al.*, 1946) are far more reliably measured with a dew-point instrument than with absorption hygrometers which, at these very low humidities, are either too slow in response (goldbeater's skin), unsuitable on account of a too high resistance (electrolytic hygrometer), or comparatively less sensitive (optical hygrometer). It is likely, however, that, with non-tarnishing base materials becoming available, the optical hygrometer, on account of its very fast response, its simple construction, its independence of temperature and its suitability as a direct-reading or recording instrument, may have some future.

§ 10. ACKNOWLEDGMENTS

My sincere thanks are due to Sir Nelson K. Johnson, Director of the Meteorological Office, Air Ministry, and to Dr. G. M. B. Dobson, F.R.S., for their great interest in this work; to the Gassiot Committee of the Royal Society for the appointment to the Mackinnon Research Studentship, 1942–1944; to Mr. A. J. Philpot, O.B.E., M.A., Director of the British Scientific Instrument Research Association for advice and technical help; to the Durham Colleges in the University of Durham, and especially to Professor F. A. Paneth, for their interest and for generous laboratory facilities.

REFERENCES

Barkas, W. W., 1942. *Trans. Faraday Soc.*, **38,** 194.

Diamond, H., Hinman, W. S., Jr. and Dunmore, F. W., 1940. *J. Res. N.B.S., Wash.*, **25,** 327.

Dobson, G. M. B., Brewer, A. W. and Cwilong, B. M., 1946. *Proc. Roy. Soc.*, A, **185,** 148.

Dunmore, F. W., 1939. *J. Res. N.B.S., Wash.*, **23,** 701.

Findeisen, W., 1937. *Wiss. Abh. Reichsanst. für Wetterdienst,* **2,** no. 11, part 2.

Glückauf, E., 1944. *Quart. J. Roy. Met. Soc.*, **70,** 293 ; 1945 a. *Ibid.*, **71,** 110 ; 1945 b. *J. Sci. Instrum.*, **22,** 34.

Sowerby, J. McG., 1944. *J. Sci. Instrum.*, **21,** 42.

A THEORY OF FLICKER NOISE IN VALVES AND IMPURITY SEMI-CONDUCTORS

By G. G. MACFARLANE,
Telecommunications Research Establishment, Ministry of Supply, Malvern

MS. received 16 *October* 1946 ; *read* 21 *February* 1947

ABSTRACT. A theory of contact noise at low frequencies is described. It is assumed that this noise is due to diffusion of clusters of mobile impurity centres on to the contact surface, as in Schottky's theory of flicker effect in valves. A cluster disappears in time due to ionization and consequent ionic conduction away from the contact surface. This diffusion-conduction process is used to derive a formula for the flicker noise which is applicable to emission from oxide-coated filaments of valves and to contacts between particles of impurity semi-conductors, as in lead sulphide photo-conductive cells and rectifiers. The spectral power density of the noise is found to depend on current j and frequency f as j^{x+1}/f^x, where $1<x<2$. Experimental results for PbS cells, oxide emitters and copper-oxide rectifiers are found to be in good agreement with this law. Carbon resistors are also found to obey this law, and it is suggested that the same mechanism of diffusion of clusters of impurity atoms on to the contact surface between crystals is responsible for the effect.

§ 1. INTRODUCTION

RECENT experimental studies of the power spectra of low-frequency noise observed in lead-sulphide photo-conductive cells, carbon resistors and copper oxide rectifiers, described by Harris, Abson and Roberts (1947), have shown that when current is flowing the power density of the noise increases rapidly as the frequency is reduced. A similar phenomenon is well known for valves with oxide-coated filaments (Johnson, 1925; Schottky, 1926) and is called *flicker effect*. This frequency-dependent noise has also been observed in silicon crystal rectifiers (Miller *et al.*, 1946) and in carbon granule microphones (Christensen and Pearson, 1936). In each case this extra noise, which we shall call flicker noise, is found to obey a law of the form

$$\overline{\Delta j^2} \propto \frac{j^m}{f^n},$$

where $\overline{\Delta j^2}$ is the mean square fluctuation of current, j the mean current and f the frequency; n ranges from 0·6 to less than 2 and m from 1·5 to 3.

Flicker noise is not the only species of noise at low frequencies. Thermal agitation noise (Johnson noise), which is independent of current, and shot noise, which is proportional to current, may also be present. In both these cases the spectral density is constant, independent of frequency. At low frequencies ($<10^3$ to 10^4 c.p.s.) flicker noise may be tens or even hundreds of times greater than Johnson or shot noise. At higher frequencies, flicker noise becomes negligible.

Consideration of the known systems in which flicker noise occurs reveals the common feature of contacts between different particles or different materials.

In most cases one of the materials at a contact is a semi-conductor. Examples are (i) the valve, in which the contact is between an excess semi-conductor, viz. an oxide of a rare earth having excess metal, and vacuum; (ii) lead-sulphide cell, which consists of a great number of micro-crystals in contact, the lead sulphide having excess lead; (iii) the copper-oxide rectifier, which consists of a deficit semi-conductor in contact with a thin insulating layer separating it from metallic copper; (iv) the carbon resistor and the carbon microphone, which consist of a multitude of carbon granules in contact. Now flicker noise occurs only when current is flowing. This suggests that it is due to variations in the conducting properties of contacts.

A theory of flicker noise in valves was described by Schottky in 1926. He ascribed the effect to fluctuations in the surface layer of foreign atoms on the cathode. The elementary event is the coming and going of separate foreign atoms or molecules, and the time constant of the event is identified with the duration of the stay of an individual atom or molecule on the surface. Each foreign atom is assumed to reduce or increase the effective work function of the surface by the same amount. Following Langmuir, Schottky accounted for this change in effective work function by supposing that a foreign atom on the surface was polarized and therefore produced an electric double layer on the surface (Schottky and Rothe, 1928). As a basis for estimating the spectral density function, Schottky assumed that the life of an adatom is controlled by a diffusion process independent of the current flowing. The correlation function for the process, which is the probability distribution of life times of adatoms on the surface, is then a pure exponential, $\exp(-qt)$, where $1/q$ is the average sojourn of an adatom on the surface. This leads to the law

$$\overline{\Delta j^2} \propto \frac{j^2}{\omega^2+q^2} \simeq \frac{j^2}{\omega^2} \text{ for } \omega \gg q,$$

where $\omega = 2\pi f$. Schottky attempted to fit this law to observations of Johnson but the fit was not very good. The experimental law was more nearly

$$\overline{\Delta j^2} \propto \frac{j^2}{\omega^{1\cdot 25}}.$$

Recently R. L. Sproull (1945) has described a surface diffusion process which satisfactorily explains the initial rapid fall in emission observed when current is suddenly drawn from a valve. At the end of his article, Sproull suggested that this decay effect might be connected with flicker effect in valves. It is our purpose to explain the connection and to use the diffusion-conduction theory to derive an expression for flicker noise. It will be shown that the theory leads to a spectral-density law that is in good agreement with experiment. Moreover it is suggested that a similar mechanism is responsible for flicker noise in other systems in which there are contacts involving impurity semi-conductors such as lead-sulphide cells, rectifiers and non-metallic resistors.

It is, however, a salient feature of our theory that the observed flicker noise arises from the diffusion of clusters of impurity centres (atoms of barium in the case of a BaO emitter) on to the emitting surface. These clusters appear only at a relatively small number of points on the emitting surface. Only in this way does it seem possible to account for the deviation of the frequency law from $1/\omega^2$.

This is in contrast to Schottky's theory, in which the fluctuations are due to the random variations in the surface density of adatoms considered as free molecules of a gas.

§ 2. THE DIFFUSION-CONDUCTION THEORY

For the present we shall restrict our discussion to the case of emission from an oxide-coated cathode. Here we have a semi-conducting mass of, say, BaO emitting electrons into a vacuum. From electron-microscopic studies it is known that the emission of electrons does not take place uniformly over the surface of the cathode but in patches. This is most apparent in the process of activation when barium atoms erupt on to the surface at scattered regions (Ahearn and Becker, 1938).* The occurrence of these eruptions means that fluctuations in the density of barium atoms on the surface will be considerably greater than the square root of the average surface density, as assumed by Schottky. The correlation function for the fluctuations can then be appreciably different from the simple exponential, as we shall show.

Let us now consider the process by which adatoms of barium appear on, and then disappear from, the surface at a brightly emitting patch when an electric field is applied.

Barium atoms diffuse on to the surface when there is a concentration gradient at the surface tending to move Ba atoms out from the interior. Barium can be transported away from the surface either by evaporation, which we shall neglect, or by ionic conduction. While barium atoms remain on the surface they are assumed to be strongly polarized and, therefore, to form a dipolar layer over the emitting patch. The strength of the double layer is proportional to the surface density of adatoms provided this is less than the density required to produce maximum emission. The effective work function is reduced by the presence of the double layer by an amount which is directly proportional to the strength of the double layer. The rate at which ionized barium atoms leave the surface is taken to be proportional to the applied electric field and, therefore, to the electronic current. The diffusion rate is proportional to the number of barium atoms that have left the surface. The diffusion-conduction equation is, therefore,

$$\frac{dN}{dt} = -\frac{\alpha}{e} j + p(N_0 - N), \qquad \ldots\ldots(1)$$

where N is the number of adatoms per cm^2 of surface at time t, j is the thermionic current density, e is the electronic charge, α is the ratio of ionic to electronic conductivities, p is the probability that a Ba atom shall move from the surface to a distance h from the surface in unit time. p is proportional to the diffusion coefficient for diffusion into the material near the surface and inversely proportional to the thickness, h, of layer in which the concentration gradient is set up. That is,

$$p = \frac{D}{h} = \frac{D_0}{h} \exp\left(-E/\boldsymbol{k}T\right), \qquad \ldots\ldots(2)$$

$$\alpha = \sigma_{\text{ion}}/\sigma_{\text{el}}. \qquad \ldots\ldots(3)$$

It is to be noted that the diffusion may not take place entirely through crystals, but to some extent along boundaries between crystals.

---

* The electron-microscope studies described by Ahearn and Becker refer to thoriated tungsten. They show that thorium comes to the surface in "eruptions" at a relatively small number of randomly located points.

The thermionic emission is related to the number of adatoms by Langmuir's equation

$$j = j_0 \,.\, e^{-a(N_0 - N)}. \qquad \ldots\ldots(4)$$

Eliminating N from (1) and (4) gives the equation of current,

$$\frac{1}{j}\frac{dj}{dt} = p \log\left(\frac{j_0}{j}\right) - \frac{(a\alpha)}{e}\, j. \qquad \ldots\ldots(5)$$

Its approximate solution is found by Sproull (*loc. cit.*) to be

$$\frac{j}{j_1} = 1 + \frac{2}{Ae^{qt} - 1}, \qquad \ldots\ldots(6)$$

where j_1 is the asymptotic value of the current for large values of t. It is found from equation (5) on putting $dj/dt = 0$. Thus

$$p \log\left(\frac{j_0}{j_1}\right) = \left(\frac{a\alpha}{e}\right) j_1. \qquad \ldots\ldots(7)$$

The decay constant q is found to be

$$q = \frac{\dfrac{2a\alpha}{e}\, j_1}{1 - \exp\left(-\dfrac{2a\alpha}{pe}\, j_1\right)}. \qquad \ldots\ldots(8)$$

This is slightly different from the expression found by Sproull and is more accurate when $\dfrac{a\alpha}{pe}\, j_1$ is small.

A is found from the boundary condition, that at time zero the number of adatoms in excess of N_1 (the number corresponding to j_1) is n. This gives

$$A = \frac{e^{an} + 1}{e^{an} - 1}, \qquad \ldots\ldots(9)$$

since

$$j(0) = j_1 e^{an}. \qquad \ldots\ldots(10)$$

We shall now use these results to derive an expression for the spectrum of flicker noise on the assumption that variations in emission arise from eruptions of barium atoms on **to** the surface at scattered points and that decay of emission from an eruption is controlled by the diffusion-conduction process described above.

§ 3. FLICKER EFFECT

Our problem essentially is to calculate the correlation function $f(\tau)$ of the noise current from an emitting patch. For the spectral power density $S(\omega)$ is the Fourier cosine transform of $f(\tau)$. Thus (Wang and Uhlenbuk, 1945)

$$S(\omega) = 4\int_0^\infty f(\tau)\cos(\omega\tau)\,d\tau. \qquad \ldots\ldots(11)$$

Now j_1 is the mean current and N_1 is the mean surface density of polarized adatoms. The increase in current due to $n + N_1$ adatoms is

$$\Delta j = j_1(e^{an} - 1). \qquad \ldots\ldots(12)$$

The correlation function of this noise current is, then,

$$\begin{aligned} f(\tau) &= <\Delta j(t) \,.\, \Delta j(t+\tau)>_{\mathrm{AV}}, \\ &= j_1{}^2 <(e^{an(t)} - 1)(e^{an(t+\tau)} - 1)>_{\mathrm{AV}}, \end{aligned} \qquad \ldots\ldots(13)$$

where the mean is to be taken over all possible values of n and t. To evaluate this mean we need to know the distribution law of lifetime of adatoms on the surface. That is to say, if we have n excess adatoms at time zero, how many of these would one expect to find after time τ? But this we have already calculated in § 3. Thus from equations (6) and (10)

$$\frac{<n(\tau)>_{\mathrm{AV}}}{n(0)}=\frac{\log\left(1+\dfrac{2}{Ae^{q\tau}-1}\right)}{\log\left(1+\dfrac{2}{A-1}\right)}. \qquad \ldots\ldots(14)$$

Therefore in equation (12) the mean value for a given n is

$$j_1^2(e^{an}-1)^2\,\frac{<e^{an(\tau)}-1>_{\mathrm{AV}}}{e^{an(0)}-1}=j_1^2(e^{an}-1)^2\,.\,\frac{A-1}{Ae^{q\tau}-1}. \qquad \ldots\ldots(15)$$

Finally, to find $f(\tau)$ we have to average over all possible values of n. This is where a knowledge of the distribution law $W(n)$ of excess adatoms enters. We have already pointed out that the standard deviation for n is not the square root of N_1 but may be many times greater. Let us therefore leave it to be found from comparision of our final expression for $S(\omega)$ with experiment and denote it by N. In the absence of any exact knowledge about $W(n)$ we shall take it to be Gaussian. Thus

$$W(n)=\frac{1}{N\sqrt{(2\pi)}}\exp\left(\frac{-n^2}{2N^2}\right). \qquad \ldots\ldots(16)$$

Then

$$f(\tau)=j_1^2\int_{-\infty}^{+\infty}\frac{A-1}{Ae^{q\tau}-1}(e^{an}-1)^2\,.\,W(n)dn, \qquad \ldots\ldots(17)$$

where A is the function of (an) given by equation (9).

Write

$$g(\tau, an)=\frac{A-1}{Ae^{q\tau}-1}, \qquad \ldots\ldots(18)$$

then equation (17) becomes

$$f(\tau)=\frac{j_1^2}{a}\int_{-\infty}^{+\infty}g(\tau, n)(e^n-1)^2\,.\,W\left(\frac{n}{a}\right)dn. \qquad \ldots\ldots(19)$$

This integral can be readily evaluated approximately since $g(\tau, n)$ is a slowly, and almost linearly, varying function of n. Thus

$$\begin{aligned}f(\tau)&=\frac{j_1^2}{a}\frac{1}{N\sqrt{(2\pi)}}\int_{-\infty}^{+\infty}g(\tau, n)(e^{2n}-2e^n+1)\exp\left(\frac{-n^2}{2a^2N^2}\right)dn,\\ &\simeq j_1^2[e^{2v}\,.\,g(\tau, 2v)-2e^{\frac{1}{2}v}\,.\,g(\tau, v)+g(\tau, 0)], \qquad \ldots\ldots(20)\end{aligned}$$

where

$$v=(aN)^2. \qquad \ldots\ldots(21)$$

The spectral power density is now obtained from equation (11). Although the Fourier transform of $g(\tau, v)$ could be evaluated, probably in series form, the following procedure leads to a more simple and illuminating formula for $S(\omega)$. It depends on the observation that the function $f(\tau)$, given by equation (20), is of the same form as

$$f_m(\tau)=c\,.\,(q\tau)^m K_m(q\tau). \qquad \ldots\ldots(22)$$

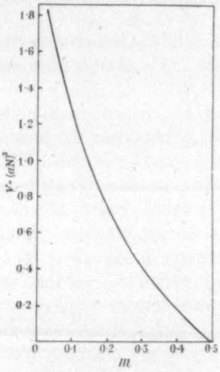
where $K_m(x)$ is the modified Bessel function of order m. Both fall off exponentially for large values of τ and rise more rapidly than $\exp(-q\tau)$ as τ tends to 0. The constant c is found by evaluating $f(0)$ and $f_m(0)$. This gives

$$c = j_1{}^2(e^{2v} - 2e^{\frac{1}{2}v} + 1)\frac{2^{1-m}}{\Gamma(m)}. \qquad \ldots\ldots(23)$$

The order m is found in terms of v by equating the integrals of $f(\tau)$ and $f_m(\tau)$ over the infinite range $0 < \tau < \infty$. This gives

$$\frac{\Gamma(\frac{1}{2})\Gamma(m+\frac{1}{2})}{\Gamma(m)} = \frac{\dfrac{2e^{2v}}{e^{2v}-1}\log\left(\dfrac{e^{2v}+1}{2}\right) - \dfrac{4e^{\frac{1}{2}v}}{e^{v}-1}\log\left(\dfrac{e^{v}+1}{2}\right) + 1}{e^{2v} - 2e^{\frac{1}{2}v} + 1}. \qquad \ldots\ldots(24)$$

m is shown as a function of $v = (aN)^2$ in the graph. It shows that as v decreases from ∞ to 0, m increases from 0 to 0·5.



Dependence of $(aN)^2$ on m.

Returning to the correlation function given by equations (22) and (23), we can now evaluate the spectral power density in closed form. Thus

$$\begin{aligned} S(\omega) &= 4\int_0^\infty f_m(\tau)\cos(\omega\tau)\,d\tau \\ &= 4j_1{}^2(e^{2v} - 2e^{\frac{1}{2}v} + 1)\frac{2^{1-m}}{\Gamma(m)}\int_0^\infty (q\tau)^m \,.\, K_m(q\tau)\cos(\omega\tau)\,d\tau \\ &= 4j_1{}^2(e^{2v} - 2e^{\frac{1}{2}v} + 1)\frac{\Gamma(\frac{1}{2})\Gamma(m+\frac{1}{2})}{\Gamma(m)}\cdot\frac{q^{2m}}{(\omega^2+q^2)^{m+\frac{1}{2}}}. \qquad \ldots\ldots(25) \end{aligned}$$

Now $S(\omega)$ given by equation (25) is the spectral power density of emission from one eruption of barium atoms on to the surface. j_1 is the average current-density of emission from patches at eruptions. But over the entire emitting surface the average fraction of area affected by eruptions is small; denote it by γ. Since eruptions are uncorrelated, the spectral power density of the noise current from the whole surface is

$$\overline{\Delta j^2} = \gamma \,.\, S(\omega). \qquad \ldots\ldots(26)$$

§4. DISCUSSION OF RESULTS

(i) The dependence of flicker noise on current and frequency is obtained from equation (25) on substituting for q from equation (8). It is convenient to normalize the current j and denote it by i, thus

$$i = \frac{2a\alpha}{pe} j_1. \qquad \ldots\ldots(27)$$

Then

$$\overline{\Delta j^2} = 4\gamma\left(\frac{pe}{2a\alpha}\right)^2 . \, p^{2m}(e^{2i} - 2e^{i} + 1)\,\frac{\Gamma(\frac{1}{2})\Gamma(m+\frac{1}{2})}{\Gamma(m)}\,\frac{i^{2m+2}}{(1-e^{-i})^{2m}}\,\frac{1}{(\omega^2+q^2)^{m+\frac{1}{2}}},$$

$$\propto \frac{i^{2m+2}}{(1-e^{-i})^{2m}}\,\frac{1}{\omega^{2m+1}} \quad \text{for } \omega \gg q, \qquad \ldots\ldots(28.1)$$

$$\propto \frac{i^{2m+2}}{\omega^{2m+1}} \quad \text{for } \omega \gg q \text{ and } i \gg 1. \qquad \ldots\ldots(28.2)$$

Therefore for large currents i and for frequencies greater than q the index of $1/\omega$ exceeds the index of i by unity. This simple relationship is therefore to be looked for in experimental results.

We have developed the above theory from the model of a thermionic emitter in which emission occurs in patches over the surface. Each patch is probably about the size of the exposed area of a crystal. This suggests that flicker noise will occur at the surface of other crystals emitting electrons, provided these crystals contain mobile polarizable impurity centres which can diffuse on to the surface and thereby alter the effective work function. It is therefore most likely to be found in multi-crystalline masses of impurity semi-conductors, of which photo-conductive lead sulphide is an example. γ is then the average fractional number of contacts at which eruptions of foreign atoms on to the contact surface occur.

In table 2 of the paper by Harris, Abson and Roberts, measured values of the current index, denoted by x, and the frequency index, y, are tabulated for lead-sulphide cells and carbon resistors. When account is taken of the scatter in points from which these index values were obtained, the agreement with the law $y = x + 1$ predicted on the above theory is satisfactory. In the case of the carbon resistors, the agreement is especially close.

(ii) The magnitude of the decay constant q and its dependence on temperature are given by equations (2) and (18). Thus

$$q = p\,\frac{i}{1-e^{-i}},$$

$$p = \frac{D_0}{h}\exp(-E/\boldsymbol{k}T),$$

and

$$i = \frac{2a\alpha}{pe} j_1. \qquad \ldots\ldots(28.3)$$

The normalized current i can also be expressed in terms of the current ratio j_0/j_1 of Sproull's experiment (1945). Thus

$$i = 2\log(j_0/j_1).$$

Therefore

$$\frac{q}{p} = \frac{2\log(j_0/j_1)}{1-(j_1/j_0)^2}.$$

Thus as j_0/j_1 increases from 1 to 100, q/p only increases from 1 to 10. Moreover, the experiments of Sproull have shown that for an oxide-coated filament (50% BaO and 50% SrO) in the temperature range 950°K. to 1100°K., j_0/j_1 is around 10. q and p are therefore of the same order of magnitude,

$$q \sim p = \frac{D_0}{h} \exp(-E/\boldsymbol{k}T).$$

From values of q at three temperatures, deducible from Sproull's results, one gets

T (°K.)	950	1020	1100
q (sec^{-1})	2500	6100	20000

The value of E is therefore about 1·23 ev. and

$$\frac{D_0}{h} \sim 8 \cdot 10^9.$$

The spectral density of the noise current should therefore flatten out to a constant value as the frequency is decreased below about 400 cycles per second at 950°K. The time constant $1/q$ is therefore of the order of hours at normal air temperatures.

In the case of other materials such as lead sulphide, the value of q depends most markedly on the value of the activation energy. This may lie between 0·5 and 3 ev. If we take $D_0/h = 10^{10}$ we find that $1/q > 20$ sec. at normal air temperature. This estimate is probably close enough to show that at normal air temperture the time constant of the diffusion-conduction process is of the order of minutes at the least. This explains why Harris, Abson and Roberts were unable to detect any flattening of the noise spectrum at frequencies as low as 1 c.p.s.

(iii) Schottky's formula for flicker noise in valves is a limiting case of our theory, viz. $N = \sqrt{N_1}$, $\gamma = 1$, and $\alpha = 0$. Since aN_1 is less than 100 and $N_1 \sim 10^{14}$, aN is a very small quantity. Then $m = 0 \cdot 5$, and one gets

$$\left.\begin{aligned} \overline{\Delta j^2} &= 4 j_1{}^2 a^2 N_1 \frac{q}{\omega^2 + q^2}, \\ &\propto j_1{}^2/\omega^2 \quad \text{for} \quad \omega \gg q. \end{aligned}\right\} \qquad \ldots\ldots (28)$$

In addition, $\alpha = 0$ means there is no ionic current and that $q = p$.

On our theory, therefore, departure of the spectral density law from $1/\omega^2$ for $\omega \gg q$ is an indication of the width of the distribution of n, whereas the departure of the current law from j^2 indicates the presence of ionic current. Schottky attempted to fit equation (28) to the spectral distribution measured by Johnson for an oxide-coated filament. The fit was not too good. The law observed was more nearly $S(\omega) \propto 1/\omega^{1 \cdot 25}$. This would correspond to $m = 0 \cdot 125$ and $(aN)^2 = 1 \cdot 11$. From equation (28.2) this flicker noise should depend on current as $j^{2 \cdot 25}$, assuming ionic current. This is in good agreement with Johnson's findings for large currents.

(iv) In order to account for the deviation of the frequency dependence of the noise power density from $1/\omega^2$ it seems necessary to postulate that impurity centres diffuse on to the emitting surface of crystals in clusters. This leads one to speculate on mechanisms that could account for the continual production and migration of clusters of atoms through a semi-conductor. A clue to a possible mechanism is to be found in the theory of photolysis of AgBr under the action of light, as proposed by Mott and Gurney (1940). To be specific let us consider the case of the BaO emitter.

The process of activation is designed to dissociate some of the BaO into free barium and to drive off the liberated oxygen. Thus in an active emitter one finds a considerable excess of barium. This free barium will most likely occur in the form of colloidal specks of barium atoms, since the process of dissociation will not be equally likely to occur at any point in the original mass of BaO, but will be more likely to commence and grow from points where the crystal lattice is distorted as at interstitial Ba ions. Moreover, specks will only be found in cracks or on the surface of crystals, for there would not be enough room for them in the body of a crystal. Now these specks of Ba will be scattered at random throughout a mass of activated BaO. However, specks which are at the exposed surface of a crystal, particularly the emitting surface of the mass of oxide, will in general be in a strong electric field, Some of the atoms in such specks will lose valency electrons, and those that remain as atoms will be strongly polarized. The loosely bound positive ions will be repelled. In fact, the speck will tend to diffuse over the entire surface of the crystal and on to adjacent crystals, forming a coating partly of Ba atoms and partly Ba ions. A selected Ba atom will alternate from the atomic to the ionic conditions. Now when the oxide is in an electric field and electrons are being drawn away from the emitting surface, ions of Ba will be conducted away from the surface. These ions will conduct through the body of crystals as well as along cracks. Eventually they will be attracted to specks of Ba. We imagine this attraction to be due to the presence of a negative charge on a speck occasioned by the capture of conduction electrons by the metallic speck, just as photo-electrons are captured by specks of Ag in the photolysis of AgBr (*vide* Mott and Gurney, p. 230). Specks of Ba will therefore grow by capture of electrons and Ba ions. Now the disintegration of a speck of Ba on the emitting surface of a crystal gives rise to a deficit of Ba on the surface. New specks of barium will therefore migrate on to the surface. This migration will be in part due to the concentration gradient near the surface and in part to conduction of specks in the electric field, since they are on the average negatively charged. Thus we have a complete cycle that accounts for the growth of specks of barium atoms, their migration to the emitting surface of the oxide, their disintegration into Ba ions at the surface, conduction of these ions away from the surface and their recombination with negatively charged specks of barium atoms. The only question outstanding is how new specks of barium are formed. Mott and Gurney encountered the same problem for specks of Ag in AgBr and explained it in terms of surface irregularities or "reifkeime". Perhaps a similar explanation is possible in this case. On the other hand we are probably safe in assuming that some of the Ba ions will conduct right through the oxide and deposit barium at the core of the emitter. Thus a layer of barium will be formed at the core and this will diffuse into the oxide along cracks and provide the nuclei of atomic barium we need.

ACKNOWLEDGMENTS

In conclusion, the author desires to express his thanks to Dr. R. A. Smith of the Telecommunications Research Establishment for his helpful criticism and encouragement, and to the Ministry of Supply for permission to publish this paper.

REFERENCES

Ahearn, A. J. and Becker, J. A., 1938. *Phys. Rev.*, **54**, 448.
Benjamin, M., Huck, R. J. and Jenkins, R. O., 1938. *Proc. Phys. Soc.*, **50**, 345.
Blewitt, J. P., 1939. *Phys. Rev.*, **55**, 713.
Christensen and Pearson, 1936. *B.S.T.J.*, **15**, 197.
Harris, E. J., Abson, W. and Roberts, W. L., 1947. (In preparation.)
Johnson, J. R., 1925. *Phys. Rev.*, **26**, 71.
Miller, Lewis, Schiff and Stephens, 1946. *Phys. Rev.*, **69**, 682.
Mott and Gurney, 1940. *Electronic Processes in Ionic Crystals* (Cambridge: The University Press).
Schottky, W., 1926. *Phys. Rev.*, **28**, 74.
Schottky and Rothe, 1928. *Handb. Experimentalphys.*, **13**, 276.
Sproull, R. L., 1945. *Phys. Rev.*, **67**, 166.
Wang, M. C. and Uhlenbuk, G. E., 1945. *Rev. Mod. Phys.*, **17**, 326.

SPONTANEOUS FLUCTUATIONS OF ELECTRICITY IN THERMIONIC VALVES UNDER RETARDING FIELD CONDITIONS

By D. K. C. MacDONALD* and R. FÜRTH†

* Military College of Science; now at Clarendon Laboratory, Oxford
† Department of Mathematical Physics, University of Edinburgh; now at Birkbeck College, University of London

MS. received 1 *October* 1946; *read* 21 *February* 1947

ABSTRACT. It follows from theoretical considerations that the shot effect in diodes should satisfy the classical formula of Schottky under extreme retarding-field conditions when the product IR of current and differential resistance of the valve reaches the constant value $\boldsymbol{k}T/e$. It appears from the results of the experiments described in the present paper, that both relations are satisfied for currents not exceeding a certain critical value I_c. A theory is presented which permits calculation of I_c for a given valve structure. Finally, it is shown how measurements of this type can be used satisfactorily for determining the cathode temperatures in diodes.

§1. INTRODUCTION

It was first pointed out by W. Schottky (1918) in a classical paper that, owing to the electronic structure of electricity, the electric current through a thermionic valve should exhibit irregular fluctuations. This phenomenon is usually called the *Shot Effect*. If it is assumed that the emission of the individual electrons from the cathode is a sequence of random events and that the transits of these electrons through the valve are independent of each other, one can derive a formula for the mean square fluctuation

$$\overline{(\delta i)^2}=\overline{(i-\bar{i})^2}$$

of the current i. Considering these fluctuations as a superposition of harmonic oscillations with frequencies f and random phases, one obtains for the part $\overline{(\delta i)_f^2}\,df$ of $\overline{(\delta i)^2}$, which is due to oscillations in a narrow region between f and $f+df$ of the spectrum

$$\overline{(\delta i)_f^2}\,df=2eI\,df, \qquad \ldots\ldots(1)$$

where e is the magnitude of the electronic charge and $I=\bar{i}$ the average current.

If the fluctuations are observed by means of an instrument (plus attached network) having a response $\phi(f)$ for an alternating current of frequency f, then the total effective fluctuation of current will be given by

$$\overline{(\delta i)^2}=2eI\int_0^\infty \phi(f)\,df=2eI\Delta f, \qquad \ldots\ldots(2)$$

where Δf is the integrated band-width of the instrument with its associated circuit.

The formulae (1) and (2) apply in the case of a valve in a state of saturation where the above-mentioned conditions are satisfied, and experiments carried out under these conditions have completely verified the theory (Williams *et al.*, 1926 and 1929). Under the ordinary working conditions of a valve, however, a negative space charge is established in the inter-electrode space which produces a variable barrier of a certain mean height for the transit of the electrons. As the height of the barrier increases with increasing current, one can easily see that the magnitude of the fluctuations will be less than the value given by formula (2), or

$$\overline{(\delta i)^2}=2eI\Gamma^2\Delta f \quad 0<\Gamma^2\leqslant 1, \qquad \ldots\ldots(3)$$

where Γ^2, the so-called ***space-charge reduction factor***, depends on the operating conditions of the valve. An extensive literature exists on the theoretical evaluation of Γ^2 (e.g. Johnson, 1925; Llewellyn, 1930; Schottky and Spenke, 1937). North and collaborators (1940–1942) in particular have carried out extensive experimental work on the measurement of the shot effect under space-charge conditions, evidencing good agreement with theory.

The above explanation of the space-charge reduction implies that $\Gamma^2=1$ when there is no variable potential barrier within the valve, i.e. when the potential either increases or decreases monotonically from cathode to anode. The first of these conditions is realized when the valve is saturated, the second when the valve is operated under retarding-field conditions, i.e. when the cold electrode has a sufficiently high negative potential with respect to the hot one. The theory therefore suggests that under true retarding-field conditions, Γ^2 should become unity once more and formula (2) should again be applicable.

It is a well-known fact that the phenomenon of emission of electrons from a hot electrode can be explained on the basis of statistical mechanics by assuming that the electrons have to overcome a fixed negative potential barrier v_c at the surface of the metal in the process of emission. Consequently there exists a probability for the escape of an electron through the surface which is proportional to $\exp\,[ev_c/\boldsymbol{k}T]$ where T is the temperature of the metal, and which is independent of the fate of the other electrons. As this expression is also quite independent of the structure of the barrier, it can equally well be applied in the case of an electronic valve under retarding-field conditions. Thus the probability of the transit of an electron from the interior of the hot electrode to the interior of the cold electrode is proportional to

$$\exp\,[e/\boldsymbol{k}T\,.\,(v_c+v_a+V)],$$

where V is the (negative) anode potential and v_a a possible potential barrier at the surface of the cold electrode. Again, the probabilities for the transit of different electrons will be independent of each other.

It appears that the conditions for the application of the formulae (1) and (2) are satisfied in both above-mentioned cases. Furthermore, the current I under true retarding-field conditions will be connected with the saturation current $\boldsymbol{J}$ by

$$I = \boldsymbol{J} \,.\, \exp\left[\frac{e(V+v_a)}{\boldsymbol{k}T}\right] = C \,.\, \exp\left[eV/\boldsymbol{k}T\right]. \qquad \ldots\ldots(4)$$

It follows immediately from (4) that the differential resistance R of the valve under these conditions satisfies the relation

$$R \equiv \frac{1}{\partial I/\partial V} = \frac{\boldsymbol{k}T}{eI}, \qquad \ldots\ldots(5)$$

showing that the product IR ought to depend on the temperature T. The relation (5) can thus be regarded as the experimental criterion for the establishment of the true retarding-field condition, and the shot effect reduction factor Γ^2 should become unity when this relation is satisfied.

An attempt to measure the shot effect in diodes under retarding-field conditions has been made before (Williams, 1936) but detailed evidence will be given below to show that the retarding region was not in fact entered on that occasion; in particular, equation (5) was not satisfied. Further, these experiments were marred by low-frequency flicker effect (Johnson, 1925; Schottky, 1926), and thus no conclusion could be drawn as to whether the shot-effect fluctuations were properly described by theory. It was therefore decided to conduct a new and more thorough investigation in order to clarify the situation. The main results have been published in a short note (Fürth and MacDonald, 1946) and a detailed account of the work is presented in what follows.

§ 2. THE FLUCTUATION MEASUREMENTS

The standard technique of fluctuation measurement that has been developed in recent years by North was employed again in this investigation. It consists in comparing the fluctuations to be measured with those generated by a diode operating under saturation condition, where equation (2) is known to be precisely satisfied (Williams, 1926 and 1929). If the same amplifier and detector are used for both measurements, the amplification factor and band width of the measuring device drop out of the final formulae and therefore need not be determined.

The circuit diagram is shown in figure 1. It consisted of the fluctuation measurement unit A which was specially designed for this investigation in the laboratory, the pre-amplifier B, the main amplifier C, a final amplifier D, the diode detector E and the measuring instrument F.

The unit A contained the test valve and the comparison valve. The test valves used were all close-spaced diodes of simple cylindrical structure with indirectly heated cathodes; the comparison valve was a directly heated tungsten-filament diode working under temperature-limited conditions. The main amplifier C was a Standard Marconi Receiver type CR.100/2; pre-amplifier and main amplifier were tuned to a frequency range around 1·5 to 2 Mc./s, to minimize interference and to avoid the "flicker effect" (Johnson, 1925; Schottky, 1926). The band width was approximately 6 kc./s. The task of the power amplifier was to provide an average input r.m.s. voltage of the fluctuations to the detector valve of the order of 50 volts; a careful examination proved that under these conditions the detector was perfectly linear within the experimental error. The measuring instrument F was a standard AVO milliammeter, the readings of which were directly proportional to the r.m.s. value of the fluctuations to be measured.

Precautions were of course taken to ensure that the effect of extraneous disturbances was reduced to a minimum throughout. Also, in addition to the meter indication, the fluctuation was displayed visually on a cathode-ray oscillograph and made audible by a loud-speaker; any occasional disturbance was then immediately evident and measurements could be suspended until the disturbance was removed.

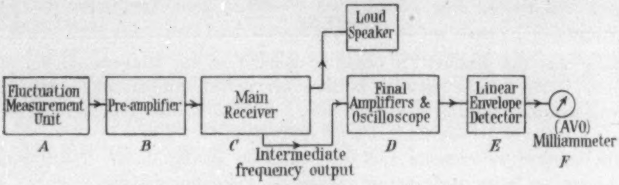

Loud Speaker
Fluctuation Measurement Unit
A
Pre-amplifier
B
Main Receiver
C
Final Amplifiers & Oscilloscope
D
Linear Envelope Detector
E
(AVO) Milliammeter
F
Intermediate frequency output

Figure 1(*a*). Schematic of fluctuation measurement layout.

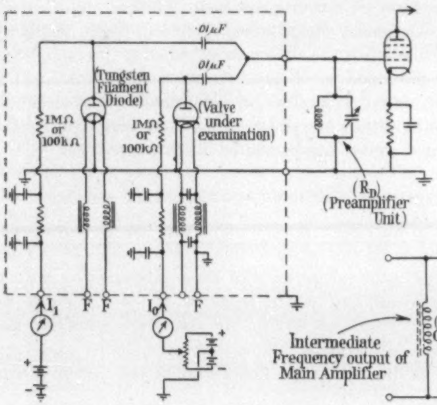

·01μF
01μF
Tungsten Filament Diode)
(Valve under examination)
1MΩ or 100kΩ
1MΩ or 100kΩ
(R_D)
(Preamplifier Unit)
I₁
F
F
I₀
F
+
-

Figure 1(*b*). Fluctuation measurement unit.

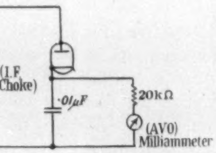

Intermediate Frequency output of Main Amplifier
(I.F. Choke)
·01μF
20kΩ
(AVO) Milliammeter

Figure 1(*c*). Envelope detector unit.

The following experimental procedure was adopted for the measurements:— First, the input to the amplifiers was short circuited and V_1, the output voltage due to inherent fluctuations introduced by the measuring device, was read. Then the short circuit was removed and the test valve alone operated at the required current I; the output voltage V_2 is then, according to (3), given by

$$V_2^2 = V_1^2 + \rho^2(2eI\Gamma^2 + 4\boldsymbol{k}T_r/R_D), \qquad \ldots\ldots(6)$$

where the third term represents the thermal fluctuations in the first circuit and ρ represents the impedance of the tuned circuit and the test valve in parallel; R_D is the dynamic resistance of that circuit and T_r its temperature. Finally the saturated tungsten diode was introduced and the current J_λ through this valve adjusted (by varying its heater current) until the output voltage V_3, which now satisfies

$$V_3^2 = V_2^2 + \rho^2 . 2eJ, \qquad \ldots\ldots(7)$$

becomes equal to

$$V_3^2 = 2V_2^2 - V_1^2. \qquad \ldots\ldots(8)$$

From (6), (7) and (8) we have for the required reduction factor:

$$\Gamma^2 = \frac{1}{I}\left(J - \frac{2\boldsymbol{k}T_r}{eR_D}\right). \qquad \ldots\ldots(9)$$

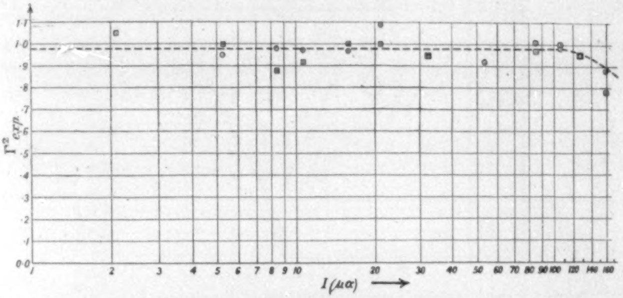


Figure 2. Experimental determination of space charge reduction factor, Γ^2_{exp}, in diode under true retarding field conditions.

○ Observed values for $I_f = 0{\cdot}3\alpha$ (mean 0·98).

⊡ Observed values for $I_f = 0{\cdot}25\alpha$ (mean 0·975).

Measurements of this kind were carried out for various valves and different cathode temperatures over a wide range of current I. Figure 2 shows graphically the results obtained on one particular diode; the other valves gave very similar results. It appears that the space-charge reduction factor is, indeed, unity up to a limiting current I_c, which in the particular case shown in figure 2 is approximately 100 μA.

§ 3. THE MEASUREMENT OF DIFFERENTIAL VALVE RESISTANCES UNDER EXTREME RETARDING-FIELD CONDITIONS

In order to check the validity of formula (5) in the true retarding field region, a method for measuring with precision the differential resistances of the valves used for the fluctuation measurements in relation to the current I had to be developed. A bridge method was employed, the circuit diagram of which is shown in figure 3. The bridge was fed from a beat frequency oscillator and frequencies between 1 kc./s. and 10 kc./s. were used. Balance was observed on a cathode-ray oscillograph after suitable amplification. A large resistance R_3 was used in series with the test valve, bridged by a large condenser C_1, to prevent slow drift of the measured current I over periods of the order of several seconds resulting from small battery variations and slow variations of emission, which latter may be expected to arise in oxide-coated cathodes when operated at low filament-temperatures.

Precautions had to be taken in order to avoid systematic errors arising from the non-linearity of the valve characteristic. As this point is of some importance for all investigations on valves in the extreme retarding-field region, the relevant mathematical treatment is given in some detail.

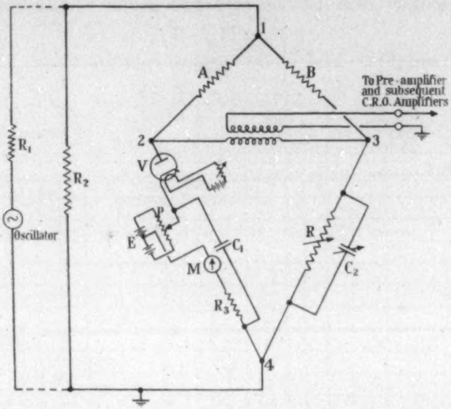


Figure 3. Detail of bridge circuit for differential valve resistance measurement.

V : Valve under examination.
R_1, R_2 : Potentiometer to provide low input voltage to bridge.
A, B : Fixed (resistive) bridge arms.
E, P : Battery and potentiometer to adjust valve current.
R : Variable (resistive) bridge arm.
C_2 : Variable condenser to balance reactive component.

According to (4) the mean current through the valve will, by the application of a small alternating voltage, $v \sin \omega t$, be changed from its original value

$$I_0 = C \,.\, \exp\,[eV_0/\boldsymbol{k}T] \qquad \ldots\ldots(10)$$

to

$$I = C \,.\, \exp\,[e/\boldsymbol{k}T(V + v \sin \omega t)], \qquad \ldots\ldots(11)$$

where V, the average anode potential, is different from V_0 because of the voltage drop across the external resistance R' due to the average increase in current. We have clearly

$$\delta \equiv V - V_0 = R'(I_0 - \bar{I}), \qquad \ldots\ldots(12)$$

and thus by means of (10) and (11), up to the second order in v,

$$\delta = -\frac{R'I_0}{4} v^2 \left(\frac{e}{\boldsymbol{k}T}\right)^2 \frac{1}{1 + R'I_0 e/\boldsymbol{k}T}. \qquad \ldots\ldots(13)$$

Now R' was chosen so as to make $R'I_0/\boldsymbol{k}T \gg 1$; accordingly (13) simplifies to

$$\delta = -\frac{v^2 e}{4\boldsymbol{k}T}, \qquad \ldots\ldots(14)$$

and the current I now becomes, to the same accuracy,

$$I = C\,.\exp\left[e/\boldsymbol{k}T\left(V_0 - \frac{\boldsymbol{v}^2 e}{4\boldsymbol{k}T}\right) + e/\boldsymbol{k}T\,.\,\boldsymbol{v}\sin\omega t\right]$$

$$\approx I_0 + \frac{I_0 e}{\boldsymbol{k}T}\left[1 - \frac{\boldsymbol{v}^2}{4}\left(\frac{e}{\boldsymbol{k}T}\right)^2\right].\,\boldsymbol{v}\sin\omega t - \frac{I_0}{4}\left(\frac{e}{\boldsymbol{k}T}\right)^2 \boldsymbol{v}^2\cos 2\omega t. \qquad \ldots\ldots(15)$$

Hence it appears that first the average current $\overline{I}$ is equal to I_0, which demonstrates the stabilizing effect of the external resistance R' mentioned above. Further, if the disappearance of the fundamental frequency $\boldsymbol{\omega}$ on the cathode-ray oscillograph trace be taken as the criterion for the balance of the bridge, the indicated differential resistance R_i of the valve is equal to the reciprocal of the coefficient of the second term in (15), i.e.

$$R_i = R\left[1 - \frac{\boldsymbol{v}^2}{4}\left(\frac{e}{\boldsymbol{k}T}\right)^2\right]^{-1} \sim R\left[1 + \frac{\boldsymbol{v}^2}{4}\left(\frac{e}{\boldsymbol{k}T}\right)^2\right], \qquad \ldots\ldots(16)$$

instead of the true value given by (5). Finally, it appears that higher harmonics of the applied oscillation are generated by the valve, the first of which is represented by the last term in (15); these harmonics are, of course, not balanced out and will appear on the oscillograph screen, thus reducing the accuracy of the measurement.

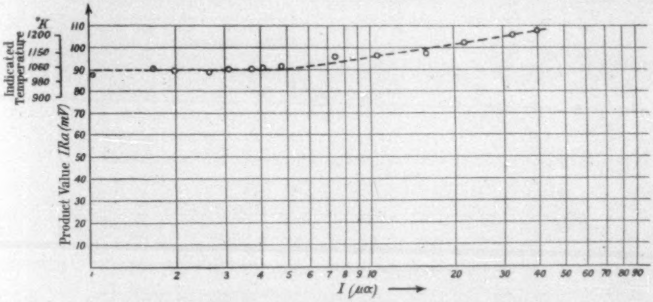


Figure 4. Experimental investigation of valve characteristic in retarding field region—Type 6H6 diode. ($I_f = 0{\cdot}3\alpha$.)

To obtain accurate results one has therefore to restrict the amplitude of the applied voltage v to sufficiently small values for $\boldsymbol{v}e/\boldsymbol{k}T$ to remain below a certain small fraction. Under the present conditions, an accuracy of 1% in the measurement of R required $\boldsymbol{v}$ to be kept in the order of magnitude of 10 mv. To confirm this result, observations were made with inputs ranging from about 2 mv. to 17 mv. and found to be entirely consistent.

In view of the small magnitude of the applied voltage, care had to be taken to ensure that stray voltage pick-up was avoided, which would vitiate the balance; and of course sufficient amplification had to be provided for the cathode-ray oscillograph.

The results obtained on all the valves examined showed clearly that below a certain value I_e of current the product IR became very accurately constant.

Examples of results on two different types of valves are shown in figures 4, 5, and 6. The results represented in figure 4 refer to a diode with a greater electrode spacing than the ones used for the experiments represented by figures 5 and 6, which were also used for the fluctuation experiments described in § 2. The measurements illustrated in figure 5 in particular were carried out on the same valve (under the same operating conditions) as that used for the

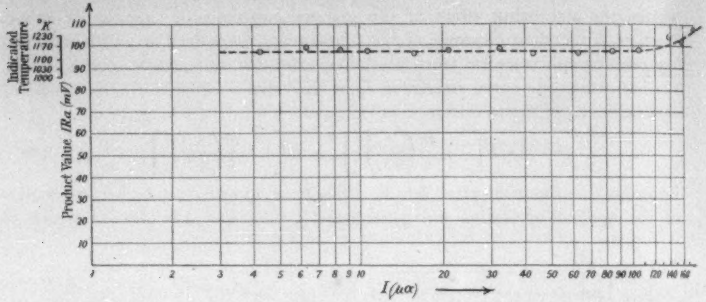


Figure 5. Examination of valve characteristic under retarding field conditions. Close-spaced diode (second model). ($I_f = 0{\cdot}3\alpha$; $V_f = 6{\cdot}9V$.)

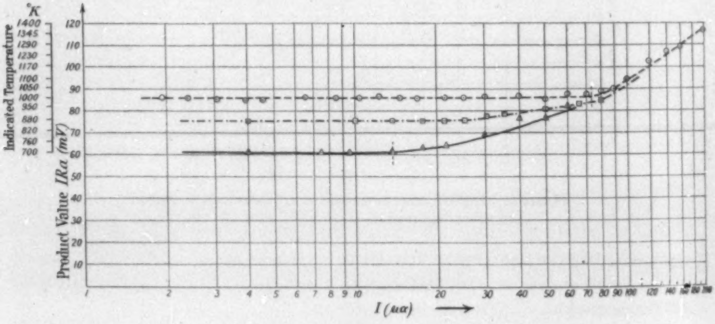


Figure 6. Examination of valve characteristic under retarding field conditions : close-spaced diode (first model).

○ Experimental points 1st series ($I_f = 0{\cdot}3\alpha$; $V_f = 6{\cdot}1V$.)

☐ ,, ,, 2nd series ($I_f = 0{\cdot}25\alpha$; $V_f = 4{\cdot}3V$.)

△ ,, ,, 3rd series ($I_f = 0{\cdot}2\alpha$; $V_f = 2{\cdot}6_5\ V$.)

measurements relating to figure 2. Figure 6 illustrates the effect of the variation of cathode temperature.

The limiting current, as indicated by these graphs, is seen to be much smaller for the widely-spaced diode than for the closely-spaced one, and (for one and the same diode) to increase with increasing cathode temperature. This is in agreement with the theoretical expectations, as will be shown in § 4. One observes also that the limiting current for the application of the fluctuation formula (2) is about the same as that for the application of formula (5) for one and the same valve, in agreement with the theoretical considerations of § 1.

§ 4. THE POTENTIAL DISTRIBUTION IN A DIODE UNDER TRUE RETARDING-FIELD CONDITIONS AND THE CRITICAL CURRENT

The results of the investigations described in §§ 2 and 3 confirm that true retarding-field conditions in a diode will prevail when the current is below a certain critical value. In the following, we attempt to derive a theoretical formula for this critical current which can be directly compared with the experimental results. For this purpose the potential distribution in a diode under retarding field conditions has first to be calculated.

Let us first consider the simplest case of a plane-parallel diode which has been treated by Langmuir (1923) and Fry (1921).

If the cathode with potential $v=0$ is at $x=0$, and the potential minimum of the space-charge barrier with $v=v_m$ is at $x=x_m$, one can, following Langmuir, introduce the dimensionless variables

$$\left.\begin{aligned} \xi &= AT^{-\frac{3}{4}}a^{-\frac{1}{2}}I^{\frac{1}{2}}(x-x_m), \\ \eta &= \frac{e}{\boldsymbol{k}T}(\boldsymbol{v}-\boldsymbol{v}_m), \end{aligned}\right\} \qquad \ldots\ldots(17)$$

where $A=4\left(\frac{\pi}{2\boldsymbol{k}}\right)^{\frac{3}{4}}m^{\frac{1}{4}}e^{-\frac{1}{2}}$, $m=$ electronic mass, and a the electrode surface area. The function $\eta(\xi)$ has been numerically computed by Langmuir, and it appears that $\xi \approx -2{\cdot}5$ for $\eta \geqslant 6$.

As there should be no potential minimum between the electrodes under the true retarding-field regime, the limiting current for this regime is given by $\boldsymbol{v}_m=V_1$, $x_m=d$, where V is the anode potential and d the distance between the electrodes. Hence for $x=0$

$$\xi_0=-AT^{-\frac{3}{4}}a^{-\frac{1}{2}}I^{\frac{1}{2}}d, \quad \eta_0=-\frac{e}{\boldsymbol{k}T}V. \qquad \ldots\ldots(18)$$

For an anode potential of $-1/2$ volt and a cathode temperature of 1000° K. one has $\eta_0 \sim 6$, and thus $\xi_0 \sim -2{\cdot}5$; hence for anode potentials larger in absolute value, the critical current I_c becomes

$$I_c \approx \left(\frac{2{\cdot}5}{A}\right)^2\frac{T^{\frac{3}{2}}a}{d^2}=\frac{(2{\cdot}5)^2\sqrt{2}}{(4\pi)^{\frac{3}{2}}}\,\frac{\boldsymbol{k}^{\frac{3}{2}}.}{e\sqrt{m}}\,\frac{I^{\frac{3}{2}}a}{d^2}=7{\cdot}5\times10^{-6}\frac{T^{\frac{3}{2}}a}{d^2}\,\mu\text{a}. \qquad \ldots\ldots(19)$$

The same problem can be treated in a much simpler way under the condition that the current I is so small that the electrons in the space-charge cloud are practically in thermodynamic equilibrium with the hot cathode of temperature T, or in other words, that the number of electrons constituting the current is small compared with the total number of electrons in the space-charge cloud.

Under this condition the (absolute) space-charge density satisfies the equation

$$\rho=\rho_0\exp\left[\frac{e\boldsymbol{v}}{\boldsymbol{k}T}\right], \qquad \ldots\ldots(20)$$

which follows directly from statistical thermodynamics. If this is combined with Poisson's equation,

$$\nabla^2\boldsymbol{v}=4\pi\rho, \qquad \ldots\ldots(21)$$

one obtains the following differential equation for $\boldsymbol{v}(x, y, z)$:

$$\nabla^2 \boldsymbol{v} = 4\pi\rho_0 \,.\, \exp\left[\frac{e\boldsymbol{v}}{\boldsymbol{k}T}\right], \qquad \ldots\ldots(22)$$

which can be solved for given boundary conditions. The density distribution $\rho(x, y, z)$ can be obtained from (20).

This problem has been extensively dealt with by Laue (1918) in a classical paper. For the linear case in question, the solution can be written in the form

$$\rho = \frac{C^2}{8\pi}\frac{e}{\boldsymbol{k}T}\frac{1}{\sin^2\left[\frac{Ce}{2\boldsymbol{k}T}(x+X)\right]}, \qquad \ldots\ldots(23)$$

where C, X are constants of integration, and C has to be real for a distribution with a maximum ρ_m of ρ at a plane $x = x_m$. For x_m and ρ_m, we have at once from (23)

$$x_m = \frac{\pi\boldsymbol{k}T}{Ce} - X, \qquad \ldots\ldots(24)$$

$$\rho_m = \frac{C^2 e.}{8\pi\boldsymbol{k}T}. \qquad \ldots\ldots(25)$$

Supposing $\boldsymbol{v} = 0$, $\rho = \rho_0$ for $x = 0$ at the cathode, we have further,

$$\rho_m = \rho_0 \sin^2\left(\frac{CeX}{2\boldsymbol{k}T}\right). \qquad \ldots\ldots(26)$$

The limit of the true retarding-field regime is given by $x_m = d$ when certainly $\rho_m \ll \rho_0$. Thus from (26),

$$CeX/2\boldsymbol{k}T \ll 1,$$

and consequently, according to (24),

$$C \approx \frac{\pi\boldsymbol{k}T}{de}; \qquad \ldots\ldots(27)$$

and hence from (25),

$$\rho_m = \frac{\pi\boldsymbol{k}T}{8ed^2}. \qquad \ldots\ldots(28)$$

Now from formulae (4) and (20) we obtain for the critical current

$$I_c = \frac{J\rho_m}{\rho_0}\exp\left[\frac{e\boldsymbol{v}_a}{\boldsymbol{k}T}\right]; \qquad \ldots\ldots(29)$$

combining this with the classical Richardson expression for the saturation current $\boldsymbol{J}$,

$$\boldsymbol{J} = \rho_0 a\sqrt{\frac{\boldsymbol{k}T}{2\pi m}}, \qquad \ldots\ldots(30)$$

one obtains finally, with the help of (28),

$$I_c = \frac{\sqrt{\pi}}{8\sqrt{2}}\frac{\boldsymbol{k}^{\frac{3}{2}}}{e\sqrt{m}}\frac{T^{\frac{3}{2}}a}{d^2}\exp\left[\frac{ev_a}{\boldsymbol{k}T}\right] = 6\times10^{-6}\frac{T^{\frac{3}{2}}a}{d^2}\exp\left[-\frac{11\,400}{T}\,|\boldsymbol{v}_a|\,(\text{volt})\right]\mu\text{a}. \qquad \ldots\ldots(31)$$

Apart from the inclusion of the additional exponential factor to allow for the existence of a possible potential barrier at the anode surface to be overcome by the electrons, we observe the agreement of (31) and (19) with a small difference in the numerical factor.

The analogous problem for the more important case of a cylindrical diode can then be treated by the same method, by solving the differential equation (22) for cylindrical symmetry, this proving extremely intractable by the first method (e.g. Wheatcroft, 1940). According to Laue, the dependence of ρ on the distance r from the centre of the cylinders is now given by

$$\rho = \frac{B^2}{2\pi e/\boldsymbol{k}T \,.\, r^2 \sin^2[B \log(r/R)]}, \qquad \text{......(32)}$$

where B, R are integration constants. It can be readily shown that this formula reduces to (23) within distances δr which are small compared with r, i.e. the plane solution can be also applied to close-spaced cylindrical diodes.

In general, the condition for a maximum ρ_m of ρ at $r = r_m$ is

$$\tan\left(B \log \frac{r_m}{R}\right) = -B, \qquad \text{......(33)}$$

leading to

$$\rho_m = \frac{1+B^2}{2\pi e/\boldsymbol{k}T \,.\, r_m^2}, \qquad \text{......(34)}$$

In the limiting case we have again the maximum appearing at the anode where $r = r_a$, and it can be shown in a manner exactly analogous to that used in the planar case that R can be put very nearly equal to r_c, the radius of the cathode cylinder. Hence from (33),

$$\frac{\tan[B \log(r_a/r_c)]}{B \log(r_a/r_c)} = -\frac{1}{\log(r_a/r_c)}, \qquad \text{......(35)}$$

From this equation B can be obtained for given values of r_a, r_c by means of a table of the function $\tan x/x$ (see e.g. Jahnke–Emde: Funktiontafeln). From (29), (30) and (34) we finally obtain for the limiting current, remembering that now $a = 2\pi l r_c$ (l : lengths of anode and cathode cylinders, $l \gg r_a$, r_c),

$$I_c = \frac{1+B^2}{\sqrt{2\pi}} \frac{\boldsymbol{k}^{\frac{3}{2}}}{e\sqrt{m}} \frac{T^{\frac{3}{2}} l r_c}{r_a{}^2} \exp\left[\frac{e v_a}{\boldsymbol{k}T}\right] = 15 \times 10^{-6}(1+B^2) \frac{T^{\frac{3}{2}} l r_c}{r_a{}^2}$$

$$\exp\left[-\frac{11\,400}{T} |v_a| \text{ (volt)}\right] \mu\text{a}. \qquad \text{......(36)}$$

When r_a/r_c is of the order of magnitude unity, the right-hand side of (35) becomes large and thus

$$B \approx \frac{\pi}{2 \log(r_a/r_c)} \gg 1,$$

so that (36) goes over into

$$I_c = \frac{\pi^{\frac{3}{2}}}{4\sqrt{2}} \frac{\boldsymbol{k}^{\frac{3}{2}}}{e\sqrt{m}} \frac{T^{\frac{3}{2}} l r_c}{r_a{}^2 \log^2(r_a/r_c)} \exp\left[\frac{e v_a}{\boldsymbol{k}T}\right]. \qquad \text{......(37)}$$

This expression is, apart from the factor r_c/r_a and the exponential factor involving v_a, identical with a formula obtained by Möller and Detels (1926) who used the Langmuir method under certain simplifying assumptions which essentially restrict its application to cases where the treatment given above is valid.

When r_a/r_c is large compared with unity, the right-hand side of (35) becomes small and

$$B \approx \frac{\pi}{\log(r_a/r_c)} \leqslant 1;$$

hence the term B^2 in (36) can be neglected.

The preceding theoretical results can now be used for calculating the expected values of I_c for the valves used in the present investigation. We restrict ourselves to those measurements where the cathode temperature was known to be the standard temperature of 1000° K. (see also § 5). The other data are collected in the following table:

Valve type	r_c (appr.) (cm.)	r_a (cm.)	l (cm.)	I_c(theor.)($v_a=0$) (μa.)	I_c (obs.) (μa.)	v_a (calc.) (v.)
6H6	0·06	0·1	0·6	24	~5	0·14
CV140	0·06	0·078	0·6	124	~100	0·02

The theoretical values of I_c, calculated from (36) for $v_a=0$ are shown in column 5 of this table, and the observed ones in column 6. One sees immediately that I_c should have a much smaller value for the relatively wide-spaced valve 6H6 than for the very close-spaced valve CV140 (we are indebted to Messrs. Ferranti, Ltd., Hollinwood, Manchester, for supplying specimens of this latter valve for this work) which agrees with experiment. This fact demonstrates clearly the advantage of using valves of the latter type for experiments on electrical fluctuations in the retarding-field region.

It is further seen that the observed values of I_c are always smaller than the theoretical ones. This fact has been observed by previous workers in this field (e.g. Möller and Detels, 1926; North, 1940) and several possible explanations have been considered by the present authors. A careful discussion of these possibilities shows, however, that none of these explanations is in fact acceptable in the present work. On the other hand, the theory presented above yields a completely natural explanation when it is assumed that v_a has a small negative value, i.e. if one assumes that not all electrons impinging upon the anode can penetrate the surface. The values of the potential barrier required for restoring agreement between theory and experiment are shown in the last column of table 1. They are of the order of magnitude of 0·1 volt and may therefore be easily accounted for by contamination of the anode surface. The effect is therefore a spurious one; it was indeed, found, to vary from valve to valve and also in the course of time in one and the same valve.

The only two workers who appear to have attempted to enter the true retarding-field region in connection with fluctuation measurements are Williams (1936) and North (1940). The former operated with currents above 25 μA. while the theoretical value (for $v_a=0$) of I_c was ~50 μA.; but in view of the facts just discussed, the actual value of I_c may well have been considerably lower, which would, of course, explain that the quantity RI had not yet reached constancy in North's experiments. Williams' fluctuation measurements were carried out under the assumption that the critical value for the current in his valve was $\sim 50\,\mu$A. An estimate, however, on the basis of his valve data gives a theoretical value (for $v_a = 0$) for I_c of the order of $1\,\mu$A. Further, in Williams' work over the current range 50–$12\,\mu$A. the factor RI diminished steadily and the indicated temperature (see § 5) was very much too high, all of which confirms that the retarding regime was not in fact entered in his work.

§ 5. DETERMINATION OF CATHODE TEMPERATURES FROM RETARDING FIELD MEASUREMENTS

It follows from equation (4) that in the retarding-field region

$$\log I = \frac{eV}{\boldsymbol{k}T} + \text{const.} \qquad \ldots\ldots(38)$$

It appears, therefore, that by plotting $\log I$ against V and measuring the slope of this curve, the cathode temperature T can be determined. Work of this kind has been undertaken in the past by Möller and Detels (1926), Heinze and Hass (1938), Demski (1929), and others. The results of these experiments agree in general with temperature estimates based on other measurements. One general disadvantage of this method, however, is the use of a logarithmic plot which can easily mask a slow "power" variation of a variable, and therefore lead to considerable errors in the measurement of cathode temperatures.

On the other hand, it follows from equation (5) that the cathode temperature can be obtained from a plot of the differential resistence R against $1/I$, or of (RI) against I:

$$T = \frac{e}{\boldsymbol{k}}(RI), \qquad \ldots\ldots(39)$$

provided that I is smaller than the critical current I_c. Evidently this method does not suffer from the source of error mentioned and should therefore give much more accurate results.

In order to demonstrate the application of this method, a scale of indicated temperatures derived from equation (39) is appended to figures 4, 5 and 6. The indicated temperatures are 1060° K. in the first case and 1150° K. in the second case, which agrees well with the standard specification of the valves and with the temperature values calculated from a formula by Widdell (1940) based on the radiation law.

The indicated temperatures in the experiments illustrated by figure 6 are 1000°, 880°, 710° K. respectively, which are in the ratios 10 : 8·8 : 7·1. According to Widdell's formula, the temperatures should be approximately in the ratios $W^{\frac{1}{4}}$ where W is the filament power (which is given at the bottom of figure 6). This leads to the ratios 10 : 8·7 : 7·3, showing excellent agreement and further illustrating the value of the method.

REFERENCES

Demski, A., 1929. *Phys. Z.*, **30**, 291.

Fry, T. C., 1921. *Phys. Rev.*, **17**, 441.

Fürth, R. and MacDonald, D. K. C., 1946. *Nature, Lond.*, **157**, 841.

Heinze, W. and Hass, W., 1938. *Z. Tech. Phys.*, **19**, 166.

Johnson, J. B., 1925. *Phys. Rev.*, **26**, 71.

Langmuir, I., 1923. *Phys. Rev.*, **21**, 419.
v. Laue, M., 1918. *Jahrb. Rad. Elekt.*, **15**, 205.
Llewellyn, F. B., 1930. *Proc. Inst. Rad. Engrs.*, **18**, 243.
Möller, H. G. and Detels, F., 1926. *Jahrb. draht. Teleg. Teleph.*, **27**, 74.
North, D. O., Harris, W. A. and Thompson, B. J., 1940. *R.C.A. Rev.*, **4**; 1941. *Ibid.*, **5**; 1942. *Ibid.*, **6**.
Schottky, W., 1918. *Ann. Phys., Lpz.*, **57**, 541.
Schottky, W., 1926. *Phys. Rev.*, **28**, 74.
Schottky, W. and Spenke, E., 1937. *Wiss. Veröff. aus Siemens-Werken*, **16**, 1.
Wheatcroft, E. L. E., 1940. *J. Inst. Elect. Engrs.*, 473.
Widdell, 1940. *R.C.A. Rev.*, **5**, 106.
Williams, F. C., 1936. *J. Inst. Elect. Engrs.*, **78**, 326.
Williams, N. H. and Huxford, W. S., 1929. *Phys. Rev.*, **33**, 773.
Williams, N. H. and Vincent, H. B., 1926. *Phys. Rev.*, **28**, 1250.

STATISTICAL ANALYSIS OF SPONTANEOUS ELECTRICAL FLUCTUATIONS

By R. FÜRTH* and D. K. C. MacDONALD†,
* Department of Mathematical Physics, Edinburgh University;
now at Birkbeck College, London University
† Military College of Science; now at the Clarendon Laboratory,
Oxford University

MS. received 9 *October* 1946; *read* 21 *February* 1947

ABSTRACT. Electrical fluctuations, generated either as "shot-effect" in a saturated diode or as "thermal fluctuations" in a tuned circuit, were produced in a receiver of high natural frequency ($\sim$100 kc./s.–1 Mc./s.) and narrow band-width ($\sim$1–6 kc./s.), and (after suitable amplification) photographically recorded by means of a cathode-ray oscillograph operating on the single-stroke system. In these circumstances the fluctuations have the character of rapid oscillations (with the natural frequency of the receiver) whose amplitude R varies slowly and irregularly in time. The study of the statistical properties of this time variation of R, as represented by the envelope of the fluctuation record, was the subject of the present investigation. In particular, the distribution function of R within a statistically stationary series of observations and the correlation between values of R separated by a finite time interval were thoroughly investigated and, in general, found to be in good agreement with the statistical theory of the phenomenon in question.

§1. INTRODUCTION

A MECHANICAL system in static equilibrium will, under close observation, always exhibit irregular fluctuations about the equilibrium position which are usually referred to as *Brownian* motion. The statistical properties of this phenomenon have been thoroughly investigated in the past.‡ They can be grouped into two classes which we, using Smoluchowski's terminology, shortly denote by the words *magnitude* and *rate* of the Brownian motion. To the first class belongs the statistical distribution of the displacements from equilibrium as observed during a long interval of time, and the averages derived from this distribution. To the second class belongs the probability function for observing a transition from one position to another during a fixed time interval and the correlation averages derived therefrom. The magnitude of the Brownian motion depends solely on the inertia of the system and the forces acting upon it, and on the temperature, but not on friction and other dissipative agencies. The rate of the Brownian motion, on the other hand, is essentially governed by these latter properties, which is especially evident in the ordinary Brownian motion of free particles in colloidal solutions.

---

‡ See those references marked with an asterisk on pp. 402 and 403.

A number of experimental investigations on various systems of the type mentioned has been carried out in which the Brownian motion of the system was continuously recorded. Statistical analysis of such records have completely confirmed every aspect of the theory. The experiments of E. Kappler (1932) on the Brownian motion of a torsion balance are the most complete of this kind; here the rate of the Brownian motion could be changed independently of its magnitude by altering the damping of the system by the variation of the pressure of the surrounding air.

The analogous phenomena in electrical circuits, which in certain limiting cases manifest themselves either as " thermal fluctuation of current " or as " shot-effect ", were first predicted in a classical paper on Brownian movement by A. Einstein (1906), and have more recently been the subject of many theoretical and experimental investigations (see e.g. Moullin, 1938). It could be shown in particular that the magnitude of the current fluctuations as expressed by the mean square of the deviations of the current from its average was in accord with theory, but no work aiming at a complete experimental statistical analysis of electrical fluctuations with regard to magnitude and rate has been published so far. This is primarily due to one difficulty inherent in this phenomenon. Whereas the Brownian motion of a mechanical system can be comparatively easily observed by optical methods (e.g. by light-beam amplification) without affecting the motion to any measurable degree, the electrical fluctuations can only be observed by means of some electrical measuring instrument which must be connected to the circuit under observation. This, however, not only alters the electrical characteristics of the circuit but also introduces a distortion of the original fluctuations owing to the mechanical properties of the measuring instrument; moreover, the mechanical Brownian motion of the measuring instrument cannot be separated from the electrical fluctuations which it should measure. These problems have been discussed, mainly by Ornstein (1927, 1938) and his collaborators (Ornstein, Burger and Taylor, 1927), in a series of brilliant theoretical and experimental investigations.

More recently, however, electronic methods of amplification and recording by means of cathode-ray oscillographs have been developed to such a perfection that electrical fluctuations can now be recorded in such a way that only the electrical properties of the circuit (including the amplifiers) are involved. Thus the way to a thorough statistical analysis of electrical fluctuations in complete analogy to the above-mentioned work on mechanical Brownian movement is opened, and some first steps in this direction have been taken by the present authors. Since on the other hand the theory of the statistics of electrical fluctuations covering both the magnitude and the speed have been developed to a high degree of perfection, for example in papers by Wang and Uhlenbeck (1945) and Rice (1945), a comparison between experiment and theory could be carried out in order to verify the underlying general principles of the theory.

In what follows, this work, on which we published a preliminary note (1946), will be described in some detail. § 2 gives a description of the experimental method and the procedure for analysing the records; § 3 contains the investigation on the magnitude of the fluctuations; § 4 some theoretical considerations on the problem of the rate of the fluctuations; and § 5 the comparison between the experiments and the results of these considerations.

§ 2. EXPERIMENTAL METHOD

In the experiments of Kappler (1932), the Brownian motion of an oscillatory system could be directly observed and recorded, as the periodic time of the torsion balance was of the order of magnitude of 30 sec. For low damping, the records have the character of harmonic oscillations, at the natural frequency of the oscillating system, whose amplitude varies irregularly in time, the speed of this variation depending on the amount of damping. The main subject of interest in this case is, therefore, not the shape of the actual oscillation record, but that of the "envelope" of the fluctuation curve which represents the variation of amplitude in time.

In order to be able to observe true electrical fluctuations it is necessary to avoid the interference of other irregular disturbances of a different nature. The most significant of these is the so-called *Flicker effect* (Johnson, 1925) which arises from some relatively gross variation of the emissive properties of valve emitters. Similar effects also occur for instance in semi-conductors. As these disturbances, which have nothing to do with the phenomenon in question, are known to vary at a comparatively slow rate, they can be effectively excluded by using oscillating circuits of high natural frequency and low damping for the generation of the electrical fluctuations. Under such circumstances, the records of the fluctuations will evidently again have the character of harmonic oscillations, the amplitude of which will vary irregularly in time at a rate determined by the overall "band-width" of the circuit plus amplifiers and recorder. It will therefore suffice to record the envelope of the fluctuation curve for fluctuations generated either by shot-effect in a diode or by thermal fluctuation in a metallic circuit, and for various band-widths.

It was decided to use a standard Marconi Communication Receiver CR/100/2 in connection with a Cossor double-beam cathode-ray oscillograph as a recording instrument. The primary fluctuations were generated in the input circuit of this receiver at various frequencies between about 100 kc./s. and 2 Mc./s. The "effective" natural frequency, however, was the customary intermediate frequency of the receiver—465 kc./s.—which is high enough to avoid the aforesaid disturbing effects. The amplification per stage of the receiver for such a frequency is also sufficiently high for the fluctuations introduced by the amplifier valves not to affect the records seriously. The band-widths provided in this receiver are normally 6 kc./s., 3 kc./s., 1·2 kc./s. and 0·3 kc./s., and were found to be very suitable for the purpose of this investigation. It was first attempted to record the detected output of the receiver which should essentially represent the amplitude variations in time. It was found, however, that considerable detail was lost in this process and, therefore, no detector was used in the final experiments, although

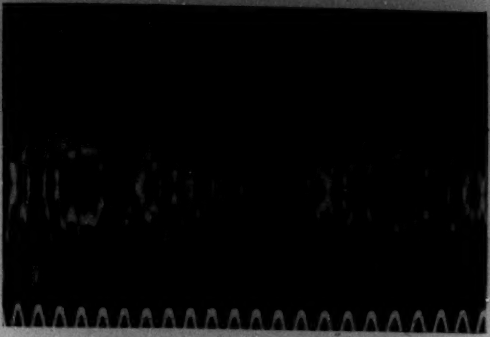


Figure 1. Photographic record of shot-fluctuation. Dominant frequency 100 kc./s., nominal band-width 6 kc./s., timing wave 1 kc./s.

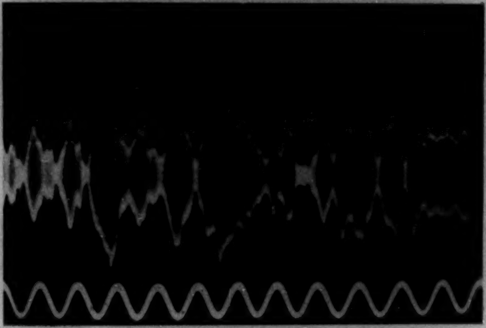


Figure 2. Photographic record of thermal fluctuation. Dominant frequency 130 kc./s., nominal band-width 1·2 kc./s., timing wave 500 c./s.

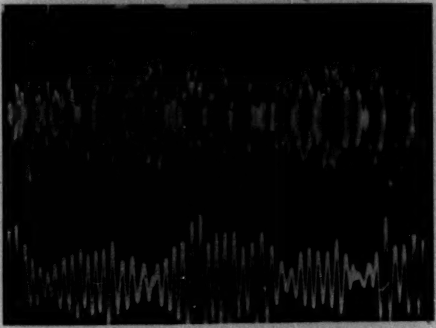


Figure 3. Upper part: record of spontaneous electrical fluctuation. Lower part: the same fluctuation, but rectified and passed through low-frequency filter with narrow response.

To face page 390

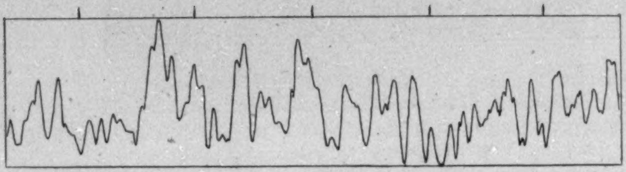


Figure 4. Traced envelope of " shot " fluctuation record. Dominant frequency 850 kc./s., nominal band-width 6 kc./s., timing marks 1 m.s.

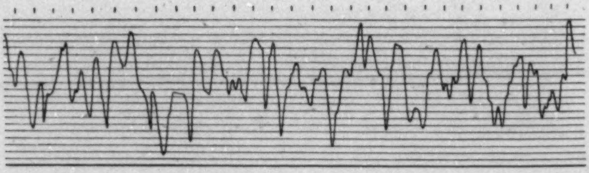


Figure 5. Traced envelope of " shot " fluctuation superposed on regular harmonic oscillation. Dominant frequency 100 kc./s., nominal band-width 1·2 kc./s., timing marks 2 m.s.

it is hoped at some later date to examine more exhaustively the design of a rectifier for such work. Owing to the rapidity of the movement of the spot on the oscillograph screen, apart from the neighbourhood of the extreme points of the oscillation, the records taken in this way show practically only the envelope curve (see figures 1 and 2).

Some experiments were undertaken, however, in order to study the character of the fluctuations obtained by letting the rectified original fluctuations pass through a low-frequency filter of limited response. A diode detector was used and the filter employed had a practically "Gaussian" response curve centred about 1000 c./s. Figure 3 gives an example of records obtained in this way together with the record of the original unrectified fluctuations. As might be expected, the phenomenon has now the character of a harmonic oscillation of 1000 c./s. frequency with amplitude slowly varying at a rate determined by the damping of the filter, and the records have a striking analogy with those obtained by Kappler for very light damping of his system.

The photographs were taken by the "single-stroke" system, incorporating a modification of the "trigger" facility already provided in the Cossor oscillograph. This method is preferable to the method of recording on a continuously moving film (as used by Kappler in his investigations) because no special camera has to be used and because the afterglow of the oscillograph screen has no ill effect in the former method. It was further found that sufficient statistical material could be obtained from one single "frame". The second beam of the oscillograph was, as a rule, used for producing a timing wave on the records, which was achieved by feeding the output from a beat-frequency oscillator on to the corresponding deflecting plates. Frequencies of 250, 500, 1000 and 2000 c./s. were used, and the timing waves are exhibited at the bottom of the records (see, for example, figures 1 and 2). In some cases the two beams were used for the simultaneous recording of two different fluctuation phenomena (see, for example, figure 3).

Photographs were taken using both a Leica and a Cossor camera; the exposure time was effectively determined by the duration of the "single-stroke" time-base and varied between about 1/10 sec. and 5/1000 sec., according to the rate of the fluctuation as governed by the band-width and the detail required. To carry out detailed analysis of the records, the original negatives were placed in a standard Leitz enlarger, and the envelopes of the curves traced on drawing paper in pencil and later traced over in Indian ink. Examples of such records are given in figures 4 and 5.

§ 3. THE MAGNITUDE OF THE FLUCTUATIONS

The first analysis was conducted to examine the distribution of R within the envelope curve $R(t)$ over a sufficiently large observation time. It was shown by Rice (1944, 1945) that the probability $p(R)dR$ for values of R within the interval $(R, R+dR)$ is given by

$$p(R)=\frac{R}{\psi}e^{-R^2/2\psi}\left(\int_0^\infty p(R)dR=1\right), \qquad \ldots\ldots(1)$$

where the constant ψ is defined by

$$\psi=\int_0^\infty W(f)df, \qquad \ldots\ldots(2)$$

$W(f)$ being the density in the "power spectrum" of the current fluctuations, and f the frequency.*

In order to check this theoretical prediction, a method was adopted which has been frequently used for similar purposes in the past (for example, Fürth, 1917; Kappler, 1932). Horizontal lines were drawn on the records at uniform vertical intervals (as seen in figures 4 and 5) and the number of intersections $q(R)$ with the trace noted. An example of a distribution obtained in this way is shown in figure 6.

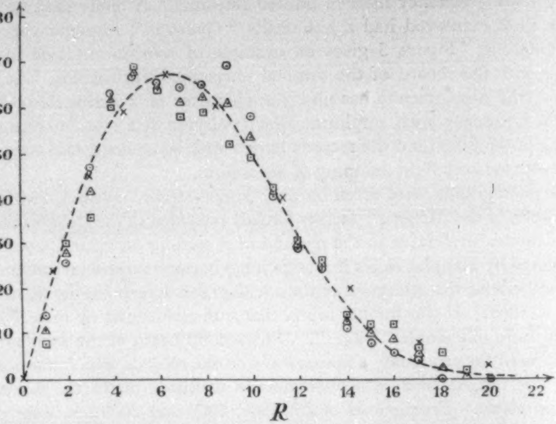


Figure 6. Statistical distribution of the amplitude, R, compared with theory.

○ Experimental points on first half ; – – – Smooth curve drawn *a priori* as best curve through these points.

□ Experimental points on second half.

△ Experimental points on whole curve to same scale.

× Computed points from $p(R)=\frac{R}{\psi}\epsilon^{-R^2/2\psi}$ scaled to agree at maximum point.

It can be shown that the distribution $q(R)$ is, apart from a constant factor, identical with the distribution $p(R)$ provided that the slope dR/dt of the curve $R(t)$ is completely uncorrelated with R everywhere†; thus the procedure of identifying

---

* Formula (1) can be derived from the general distribution formula for the fluctuation current I, that is, the probability $p(I)dI$ that $I(t)$ lies between I and $I+dI$ which, according to Rice, is

$$p(I)dI=\frac{1}{\sqrt{2\pi\psi}}e^{-I^2/2\psi}\,dI.$$

From this it follows that the mean square of the current is equal to $\overline{I^2}=\psi$. In the particular case considered here, where the band-width of the system is small compared with the mid-frequency, the current fluctuation can be regarded as the superposition of sine and cosine oscillations (at the dominant frequency of the system), whose amplitudes vary irregularly and independently in time according to the above formula.

† Call δt_1, δt_2, ... the time intervals during which $R(t)$ is found in the narrow range between R and $R+\delta R$ within the total observation time T. Then by definition

$$p(R)\delta R=(\delta t_1+\delta t_2+\ \ldots\ .)/T=\left(\frac{\delta t_1}{\delta R}+\frac{\delta t_2}{\delta R}+\ \ldots\ .\right)\frac{\delta R}{T},$$

from which follows

$$p(R)=\frac{1}{T}\left(\overline{\frac{dt}{dR}}\right)q(R).$$

If, now, dR/dt is uncorrelated with R, the quantity $\overline{(dt/dR)}$ is independent of R and thus $p(R)$ is proportional to $q(R)$, as stated in the text.

$q(R)$ with $p(R)$ is justified under this condition but completely unjustified, e.g., for a periodic curve $R(t)$. In order to make sure about this point in the present problem a " detailed " analysis was also carried out in some cases by subdividing the t-axis into narrow intervals (as seen in figure 4) and determining the distribution of the corresponding values of R. The result was essentially the same as that obtained by the first method, which was accordingly generally adopted in view of its greater simplicity.

Before going further, another necessary check had to be made to ascertain that the fluctuation was statistically stationary over the whole observation time and that a possible burst of local interference had not vitiated the record. To this end the records were frequently divided into two halves and the distribution of each half studied separately to see whether they fitted to each other as demonstrated by figure 8. Any record which was found not to satisfy this test was discarded.

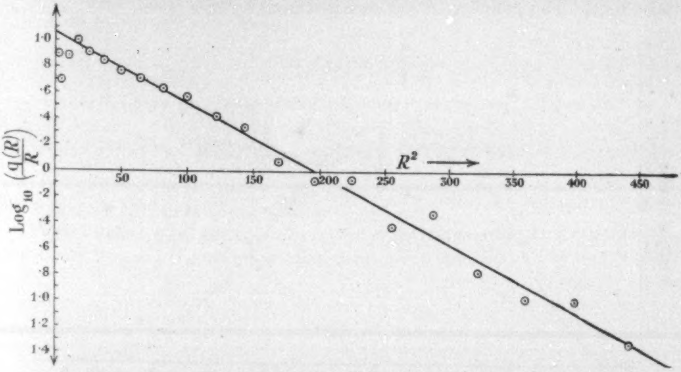


Figure 7. Logarithmic representation of the statistical distribution of amplitudes R.

The simplest way to prove whether the observed distribution agrees with the theoretical formula (1) is to draw a smooth curve through the observation points and fit the available constants so that the observed maximum of this curve coincides with the maximum of the function (1). This procedure is illustrated in figure 6, and one sees that there is no systematic deviation of the observed from the theoretical points. A more stringent test can be applied by plotting $\log[p(R)/R]$ against R^2, which, according to (1), should give points lying on a straight line with slope $-1/2\psi$. An example of such a plot is presented in figure 7, which shows that this is indeed the case, leading to the value $\psi = 40$ for this particular set of observations.

The average moments and the most probable value R_m of R, i.e. that value of R for which $p(R)$ is a maximum, can be easily derived from formula (1); one has, for instance,

$$\left.\begin{aligned} \bar{R} &= \sqrt{\pi\psi/2}, & \overline{R^3} &= 3\psi\sqrt{\pi\psi/2}, \\ \overline{R^2} &= 2\psi, & R_m &= \sqrt{\psi}, \end{aligned}\right\} \qquad \ldots\ldots(3)$$

whence

$$\frac{(\bar{R})^2}{\overline{R^2}} = \frac{\pi}{\psi}, \qquad \frac{\overline{R^3}}{\bar{R}\,.\,\overline{R^2}} = \frac{3}{2}, \qquad \frac{R_m^2}{\overline{R^2}} = \frac{1}{2}. \qquad \ldots\ldots(4)$$

By direct computation from the observed values of R, the following "experimental" values were obtained for the above-mentioned set of observations :

$$\overline{R^2}\,(\text{exp}) = 81, \qquad R_m\,(\text{exp}) \sim 7,$$

which compare well with the theoretical values derived from (3):

$$\overline{R^2}\,(\text{theor}) = 80, \qquad R_m\,(\text{theor}) = 6{\cdot}3.$$

Another observation series yielded the following experimental values for the ratios of the moments:

$$(\overline{R})^2/\overline{R^2}\,(\text{exp}) = 0{\cdot}75, \qquad \overline{R^3}/\overline{R}\,.\,\overline{R^2}\,(\text{exp}) = 1{\cdot}7,$$

which again agree well with the theoretical values 0·785 and 1·5.

A further interesting problem which belongs to the same class of magnitude fluctuation phenomena is the distribution of the frequencies with which maxima of different heights occur in the envelope curve during a sufficiently long observation time. This problem has also been discussed by Rice (1944, 1945), and his analysis shows that this distribution function is solely determined by the electrical properties of the network concerned, as was to be expected. The experimental data for the derivation of the distribution function of the maxima were therefore collected for a number of "frames" and a typical example is shown in figure 8. The spread of the observational points is naturally very considerable owing to the limited number of maxima within a frame. No detailed computation was therefore carried out, but the general shape of the smooth curve drawn (rather arbitrarily) through the observational points agrees with the ones derived by Rice from his theory. Much longer series of observations would have to be carried out were a detailed check of the theory to be attempted.

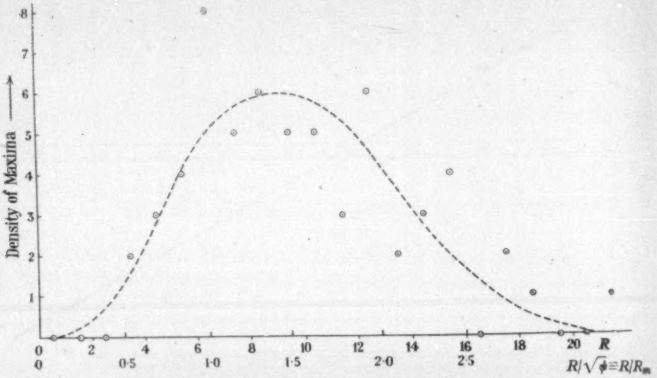


Figure 8. Statistical distribution of the heights of the maxima in the envelope records $R(t)$.

○ Experimental points.
- - - - Smooth curve through experimental points (extremely small maxima are omitted).

A special series of experiments was made in order to study the effect of the superposition of a regular harmonic oscillation on the irregular fluctuations.

For this purpose a regular "signal" (from a signal generator) was injected at the receiver input (tuned to the same frequency) and the fluctuations were recorded in the usual way.

The theory of this phenomenon is also presented in Rice's papers. He obtains for the distribution function $p(R)$:

$$p(R)=\frac{R}{\psi}\exp\left[-\frac{R^2+S^2}{2\psi}\right].\,I_0\left(\frac{RS}{\psi}\right), \qquad \ldots\ldots(5)$$

where S is the amplitude of the signal as it would appear on the record when no fluctuations were present, and where I_0 is the Bessel function of zero order with imaginary argument. From (5) follows, for the mean square of R,

$$\overline{R^2}=S^2+2\psi \qquad \ldots\ldots(6)$$

as a direct consequence of the fact that the regular oscillations and the random oscillations are not correlated. When S becomes "large", one can replace $I_0(z)$ by its asymptotic expression:

$$\lim_{z\to\infty} I_0(z)\approx\frac{e^z}{\sqrt{2\pi z}} \qquad \ldots\ldots(7)$$

(apart from the very beginning of the distribution curve), and (5) goes over into

$$p(R)\sim\frac{1}{\psi}\sqrt{\frac{R}{2\pi S}}\exp\left[-\frac{(R-S)^2}{2\psi}\right], \qquad \ldots\ldots(8)$$

which shows that the fluctuation amplitude is almost normally distributed about the amplitude of the regular oscillation.

Figure 5 is a typical example of a record of $R(t)$ obtained in these experiments, and figure 9 shows the corresponding distribution function. The computation of the mean square of R gave $\overline{R^2}=239{\cdot}4$ and the abscissa of the maximum of the curve gives $S=14{\cdot}3$. Thus from (6), $2\psi=35{\cdot}4$ and $a=S/\sqrt{\psi}=3{\cdot}4$. Rice has computed the function $p(R)$ for various values of a from (5) and the result is

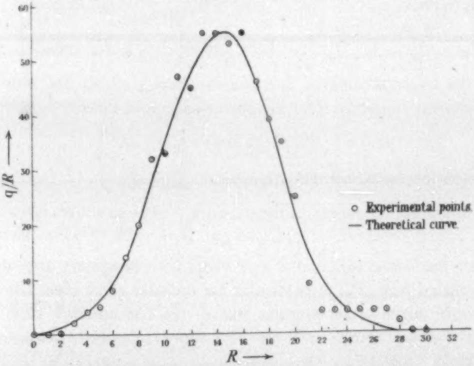


Figure 9. Statistical distribution of the amplitudes R of a fluctuation superposed on a regular harmonic oscillation, compared with theory.

presented in his paper in a set of curves. As these only vary slowly in shape as the parameter a varies, the curve relating to $a=3$ (which is nearest to the actual value 3·4) was taken and reduced to scale so as to make the maxima of the experimental and the theoretical curves coincide. This theoretical curve is shown in figure 11 and is seen to fit the observational points very satisfactorily.

§ 4. SOME THEORETICAL CONSIDERATIONS REGARDING THE FLUCTUATION RATE

We now discuss the second class of fluctuation phenomena which, as indicated in § 1, are connected with the rate at which the fluctuations occur or, in other words, with the degree of correlation between fluctuations separated by a finite time interval. If it is assumed that the primary effects which give rise to the fluctuations are completely uncorrelated, this correlation will be entirely due to the electrical characteristics of the network used. If, on the other hand, these primary effects are already correlated to some degree, the observed rate would also be affected by this primary correlation.

The theory of the rate phenomena under the assumption of complete randomness of the primary effects has been developed by Rice in particular. R. E. Burgess has also contributed to this field in unpublished work. Some further aspects of the theory have been developed by the present authors. The object of the analysis of the fluctuations with respect to rate was to check these theoretical relations and to see whether there are any indications that the correlation of the primary effects might have to be taken into account.

The basic formula to start from is that for the probability $p(R_1, R_2)dR_1dR_2$, of observing a value between R_1 and R_1+dR_1 at time t and a value between R_2 and R_2+dR_2 at time $t+\tau$ within a statistically stationary series of observations. This is, according to Rice,

$$p(R_1R_2)=\frac{R_1R_2}{A}\,.\,I_0\left[\frac{R_1R_2}{A}(\mu^2+\lambda^2)^{\frac{1}{2}}\right]\exp\left[-\frac{\psi}{2A}(R_1{}^2+R_2{}^2)\right], \quad \ldots\ldots(9)$$

where ψ is the quantity defined by (2) and where μ, λ are the sine and cosine Fourier transforms respectively of the power-spectrum density function $W(f)$:

$$\left.\begin{aligned}\mu(\tau)&=\int_0^\infty W(f)\cos 2\pi(f-f_0)\tau\,.\,df,\\ \lambda(\tau)&=\int_0^\infty W(f)\sin 2\pi(f-f_0)\tau\,.\,df,\end{aligned}\right\} \quad \ldots\ldots(10)$$

f_0 being the dominant frequency of the network. A is an abbreviation:

$$A=\psi^2-\mu^2-\lambda^2, \quad \ldots\ldots(11)$$

and I_0 is again the Bessel function of zero order with imaginary argument.

Rice calculated $p(R_1, R_2)$ in particular for the case of an ideal band-pass filter with a response curve of rectangular shape. In the present case, where the response curve of the receiver used was very approximately "Gaussian" in shape, we can set

$$W(f)=\frac{\psi}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(f-f_0)^2}{2\sigma^2}\right], \quad \ldots\ldots(12)$$

which satisfies (2) when $\sigma \ll f_0$. Under the same condition one obtains easily from (10), (11), (12):

$$\mu=\psi z, \qquad \lambda=0, \qquad A=\psi^2(1-z^2), \qquad \text{where } z=e^{-2\sigma^2\pi^2\tau^2}. \quad \ldots\ldots(13)$$

Introducing the quantity

$$v_\tau = R_2 - R_1, \qquad \ldots\ldots(14)$$

we can now define the probability $P(v_\tau)dv_\tau$ for observing a fixed difference between v_τ and $v_\tau + dv_\tau$ of R over a given time interval τ, irrespective of R_1 within the series of observations:

$$P(v_\tau) = \int_0^\infty p(R_1, R_1 + v_\tau)dR_1 \qquad (v_\tau > 0) \qquad \ldots\ldots(15)$$

which, of course, will be independent of time. As (9) is symmetrical in R_1, R_2, obviously $P(-v_\tau) = P(v_\tau)$.

The average value Δ_τ of $|v_\tau|$ is defined by

$$\Delta_\tau = \overline{|v_\tau|} = 2\int_0^\infty v_\tau \,.\, P(v_\tau)dv_\tau, \qquad \ldots\ldots(16)$$

for which in analogous problems of Brownian motion the term "after-effect" has been introduced by Smoluchowski (1923). The quantity

$$\epsilon = \lim_{\tau = 0} (\Delta_\tau / \tau)$$

is a convenient measure for the overall speed of the fluctuations. It may be remarked, however, that the fluctuations will still take place at a finite rate even if this quantity is zero or infinite.

We first consider the case of τ very small. Under this condition we obtain from (13) approximately

$$\mu = \psi(1 - 2\pi^2\sigma^2\tau^2), \qquad A = 4\pi^2\sigma^2\tau^2\psi^2. \qquad \ldots\ldots(17)$$

As now the argument of I_0 in (9) becomes very large except for small values of R_1R_2, we can again use the asymptotic expression (7); hence

$$\lim_{\tau \to 0} p(R_1R_2) = \frac{\sqrt{R_1R_2}}{(2\pi\psi)^{3/2}\sigma\tau} \exp\left[-\frac{R_1R_2}{2\psi} - \frac{(R_1 - R_2)^2}{8\pi^2\psi\sigma^2\tau^2}\right]. \qquad \ldots\ldots(18)$$

From (14), (15) and (18) we obtain by simple calculation

$$\lim_{\tau = 0} P(v_\tau)dv_\tau = \frac{1}{\sqrt{\pi}} e^{-x^2} dx \quad \left(\int_{-\infty}^{+\infty} P(v_\tau)dv_\tau = 1\right), \qquad \ldots\ldots(19)$$

where
$$x = \frac{v_\tau}{2\pi\sigma\tau\sqrt{2\psi}}, \qquad \ldots\ldots(20)$$

which shows that the quantity v_τ is normally distributed about zero, as was to be expected.

From (16) we now get

$$\lim_{\tau \to 0} \Delta_\tau = 2\sqrt{2\pi\psi} \,.\, \sigma\tau, \qquad \ldots\ldots(21)$$

which shows that the "speed" ϵ of the fluctuations is finite and equal to

$$\epsilon = \sqrt{2\pi\psi} \,.\, \sigma.$$

We now turn to the opposite limiting case of τ very large. Here we have from (13) $\mu = 0$, $A = \psi^2$, and since $I_0(0) = 1$, formula (9) becomes

$$\lim_{\tau \to \infty} p(R_1R_2) = \frac{R_1R_2}{\psi^2} e^{-\frac{R_1^2 + R_2}{2\psi}}, \qquad \ldots\ldots(22)$$

which also follows immediately from (1), as in this case there is evidently no correlation between the fluctuations of R_1 and R_2.

Similarly, as above, we obtain now

$$\lim_{\tau\to\infty} P(v_\tau)\,dv_\tau = e^{-y^2}\{ye^{-y^2}+\sqrt{\pi}(\tfrac{1}{2}-y^2)(1-\operatorname{erf}y)\}\,dy, \qquad \ldots\ldots(23)$$

where

$$y=\frac{|v_\tau|}{2\sqrt{\psi}} \qquad \ldots\ldots(24)$$

and

$$\operatorname{erf}y \equiv \frac{2}{\sqrt{\pi}}\int_0^y e^{-u^2}\,du \qquad \ldots\ldots(25)$$

is the well known error function.

In the general case* of an arbitrary τ it is advantageous to introduce instead of R_1 the new variable

$$t=\frac{R_1(R_1+v_\tau)}{\psi(1-z^2)} \qquad \ldots\ldots(26)$$

from (9), (13), (15), and one now obtains easily

$$P(v_\tau)=(1-z^2)\,.\exp\left[-\frac{v_\tau^{\,2}}{2\psi(1-z^2)}\right]\int_0^\infty \frac{te^{-t}I_0(zt)}{\sqrt{v_\tau^{\,2}+4\psi(1-z^2)t}}\,dt. \qquad \ldots\ldots(27)$$

Using the identity

$$\frac{1}{\sqrt{\pi}}\int_0^\infty \xi^{-\frac{1}{2}}e^{-\alpha\xi}\,d\xi \equiv \frac{1}{\sqrt{\alpha}} \qquad \ldots\ldots(28)$$

we can transform (27) into

$$P(v_\tau)=\frac{\sqrt{1-z^2}}{2\sqrt{\pi\psi}}\cdot\exp\left[-\frac{v_\tau^{\,2}}{2\psi(1-z^2)}\right]\cdot\int_0^\infty \xi^{-\frac{1}{2}}e^{-\frac{v_\tau^{\,2}\xi}{4\psi(1-z^2)}}\,d\xi\int_0^\infty te^{-t(1+\xi)}I_0(zt)\,dt. \qquad \ldots\ldots(29)$$

Now it may be shown by Hankel's formula (Whittaker and Watson, 1935) that

$$\int_0^\infty t\,.\,e^{-t(1+\xi)}\,I_0(zt)\,dt=\frac{1+\xi}{\xi^{\frac{1}{2}}[(1+\xi)^2-z^2]^{3/2}}. \qquad \ldots\ldots(30)$$

Thus finally,

$$P(v_\tau)=\frac{\sqrt{1-z^2}}{2\sqrt{\pi\psi}}\exp\left[-\frac{v_\tau^{\,2}}{2\psi(1-z^2)}\right]\int_0^\infty \frac{1+\xi}{\xi^{\frac{1}{2}}[(1+\xi)^2-z^2]^{3/2}}\,e^{\frac{v_\tau^{\,2}\xi}{4\psi(1-z^2)}}\,d\xi. \qquad \ldots\ldots(31)$$

From (16) and (31) we get for the after-effect for arbitrary τ

$$\Delta_\tau=2\sqrt{\frac{\psi(1-z^2)^3}{\pi}}\int_0^\infty \frac{1+\xi}{\xi^{\frac{1}{2}}(2+\xi)[(1+\xi)^2-z^2]^{3/2}}\,d\xi \qquad (z=e^{-2\pi^2\sigma^2\tau^2}). \qquad \ldots\ldots(32)$$

In particular for $\tau\to\infty$ (the case considered above) one has

$$\Delta_\infty=2\sqrt{\frac{\psi}{\pi}}\int_0^\infty \frac{d\xi}{\xi^{\frac{1}{2}}(1+\xi)^2(2+\xi)}=(\sqrt{2}-1)\sqrt{\pi\psi}=0{\cdot}73\sqrt{\psi}. \qquad \ldots\ldots(33)$$

This differs from $\bar{R}$ (formula (3)) only by a numerical factor because of the complete lack of correlation for long time intervals.

For values of τ not too small the quantity z is small and the integral in (32) can be expanded into a power series with respect to z:

$$\Delta_\tau=(\sqrt{2}-1)\sqrt{\pi\psi(1-z^2)^3}\{1+1{\cdot}035z^2+1{\cdot}046z^4+1{\cdot}046z^6+\ldots\},$$

---

* We have to thank Dr. A. Erdélyi for his valuable advice on the solution of this problem.

which is approximately equal to

$$\Delta_\tau = (\sqrt{2}-1)\sqrt{\pi\psi(1-z^2)} = (\sqrt{2}-1)\sqrt{\pi\psi}\sqrt{1-e^{-4\pi^2\sigma^2\tau^2}}. \quad \ldots\ldots(34)$$

This, of course, goes over into (33) for $\tau=\infty$. For small values of τ, on the other hand, the expression on the right-hand side of (34) becomes

$$(\sqrt{2}-1)\sqrt{\pi\psi} \,.\, 2\pi\sigma\tau = 0{\cdot}84 \,.\, 2\sqrt{2\pi\psi} \,.\, \sigma\tau,$$

which differs from the correct expression (21) by the numerical factor 0·84. Thus it is seen that formula (34) represents the function Δ_τ with very good approximation over almost the whole range of τ from 0 to ∞.

§5. COMPARISON BETWEEN THE THEORY OF THE FLUCTUATION RATE AND EXPERIMENT

The theoretical results of §4 were put to the test by subdividing the t-axes of the fluctuation records into equal time intervals of length τ and reading off the differences v_τ of the successive ordinates. Thus the frequency distribution of the different values of v_τ within a record and the average Δ_τ can be obtained and compared with the theoretical formulae.

First the limiting case of small τ was studied on a number of fluctuation records, using time intervals corresponding to a distance of 1 mm. on the records, which were just large enough to make the measurements of the ordinates reasonably accurate. The distribution of v_τ for two records, one of "shot" fluctuations, the other of "thermal" fluctuations, is displayed in figure 10, together with the theoretical Gauss distribution curve (19) drawn in such a way as to give the best fit of the observed points. This shows that formula (19) is indeed satisfied. Moreover, the value of the constant $C=v_\tau/x$ can now be determined for each record and compared with its theoretical value $2\pi\sigma\tau\sqrt{2\psi}$ according to (20). τ is readily determined from the timing wave on the record, and $\psi=\overline{R^2}/2$ (see formula (3)) is found by the method described under §3. σ can be determined from the frequency-response curve of the receiver, observed under the same operating condition under which the fluctuation record is obtained, and by comparing this curve with formula (12) by numerical integration. The results for three records, including the two mentioned above, are collected in table 1 and show a very satisfactory agreement between the theoretical and experimental values of C.

Table 1

Record	$\tau\times10^5$ sec.	$\sqrt{\overline{R^2}}$	$\sigma\times10^{-3}$ sec.$^{-1}$	C (theor.)	C (exp.)	$\frac{C\text{ (exp.)}}{C\text{ (theor.)}}$
Shot fluctuation (4—J)	2·2	9·0	1·95	2·42	2·77	1·14
Thermal fluctuation (7—J)	2·24	9·6	2·08	2·76	2·76	1·00
Thermal fluctuation (16—C)	4·5	9·4	1·19	3·14	3·26	1·04

We now turn to the experimental results for large intervals. To examine the frequency distribution of v_τ, the differences of the ordinates R at 30 mm. apart were employed for the first record mentioned above. As Δ_τ was found to reach a stationary value for $\tau > 10$ mm. say (see figure 12) in this case, the separation employed should correspond well to the theoretical case of $\tau \to \infty$. The observed distribution for this record is shown in figure 14. From the experimental value $\overline{R^2} = 81$ for this particular series we obtain $2\sqrt{\psi} = 12{\cdot}7$. Using this scale factor for the abscissae according to (24), the observed points should lie on the theoretical curve (23). This curve is also shown in figure 11, and the observed points are seen to fit the theoretical curve satisfactorily.

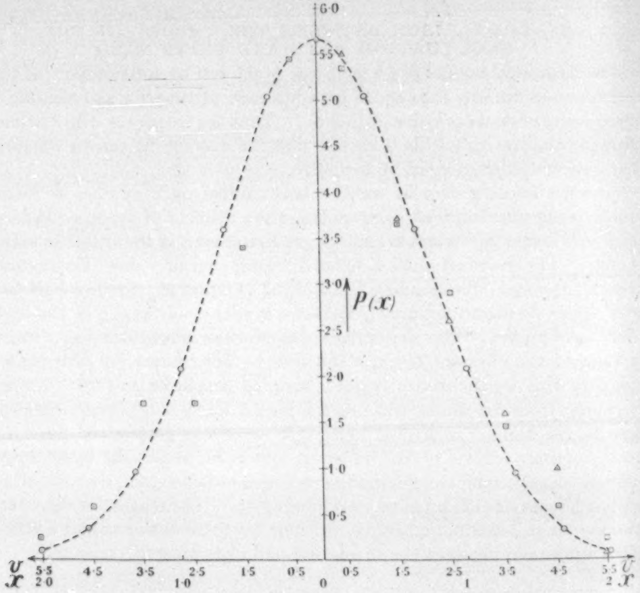


Figure 10. Statistical distribution of v_τ for small τ, compared with theory.

○ - - - } $P(x) = \frac{1}{\sqrt{\pi}} \epsilon^{-x^2}$.

⊡ Experimental points for shot fluctuation.

△ Experimental points for thermal fluctuation.

In addition, the limiting values of Δ_∞ (the average of $|v_\infty|$) were estimated from a number of records; at the same time the theoretical values of Δ_∞ were calculated from (33). These are exhibited in table 2, and again agreement between theory and experiment is established.

No check of the general distribution formula (31) was attempted, but formula (34) for Δ_τ was put to the test by computing the values of $\Delta_{\tau n}$ for the basic time interval τ (corresponding to 1 mm. spacing on the records) and its integer multiples

Table 2

Record	$\sqrt{\overline{R^2}}$	Δ_∞ (theor.)	Δ_∞ (exp.)	$\dfrac{\Delta_\infty(\text{exp.})}{\Delta_\infty(\text{theor.})}$
Shot fluctuation (4—J)	9·0	4·66	4·6	0·98
Shot fluctuation (6—J)	10·07	5·21	5·9	1·13
Thermal fluctuation (7—J)	9·06	4·98	4·8	0·96
Thermal fluctuation (7—J)	12·93	6·71	7·2	1·07

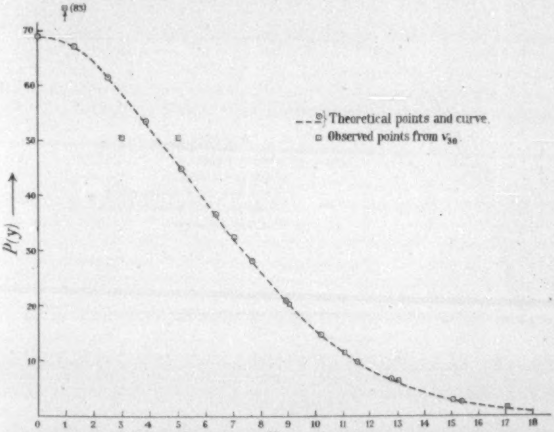


Figure 11. Statistical distribution of v_τ for large τ, compared with theory.

$n\tau$, again for a number of shot and thermal fluctuation records. In order to facilitate the comparison with theory the reduced after-effect factors

$$\delta_n = \frac{\Delta_{\tau n}}{\Delta_\infty} \qquad \ldots\ldots(35)$$

were introduced, which according to (33) and (34) should satisfy the formula

$$\delta_n = \sqrt{1 - e^{-4\pi^2\sigma^2\tau^2n^2}}. \qquad \ldots\ldots(36)$$

An example of the results is shown in figure 12. The broken line represents the function (36) for the values of the constants given in table 1. This function is inaccurate for small values of *n*, as explained in § 4. The beginning of the correct curve, on the other hand, can be calculated from (12). By joining this with the former curve, the correct theoretical curve is obtained drawn as a continuous line in figure 12. The observed values of δ_n, i.e. the values $\delta_n(\text{obs.}) = \Delta_{rn}(\text{obs.})/\Delta_\infty(\text{obs.})$ are also indicated in the diagram and connected by the dotted lines.

The agreement between experiment and theory appears to be satisfactory. It appears therefore that the correlations exhibited by the results of the present investigation are entirely due to the macroscopic electrical characteristics of the network used, as supposed in the theory. This was, of course, to be expected because of the limited band-widths which had to be employed. It would be valuable indeed to extend these investigations to much greater band-widths, which would perhaps reveal some differences in the statistical character of the fluctuations from shot and thermal sources due to differences in the correlation between the primary events responsible for these effects. Definite conclusions on this problem will therefore have to be postponed until further and more extensive experimental material on this subject becomes available.

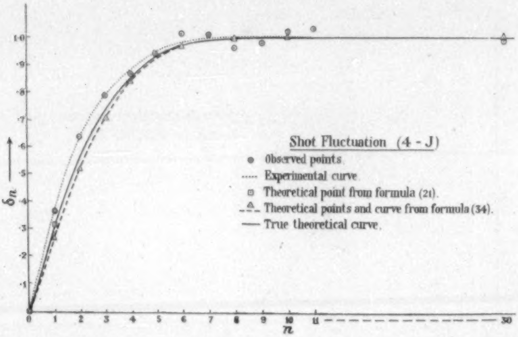


Figure 12. Statistical distribution of δ_n for shot fluctuation, compared with theory.

REFERENCES

BARNES, R. BOWLING and SILVERMAN, S. 1934. *Rev. Mod. Phys.*, **6**, 162.
*CHANDRASEKHAR, S., 1943. *Rev. Mod. Phys.*, **15**, 1.
*DE HAAS-LORENTZ, Mrs. G. L., 1913. *Die Brownsche Bewegung und einige verwandte Erscheinungen* (Braunschweig).
EINSTEIN, A., 1906. *Ann. Phys., Lpz.*, **19**, 380.
*EINSTEIN, A., 1922. "Untersuchungen über die Theorie der Brownschen Bewegung" (Leipzig: Ostwald's *Klassiker der exakten Naturwissenschaften*, Nr. 199); 1926. *Investigations on the theory of the Brownian Movement* (London: Methuen).
FÜRTH, R., 1917. *Ann. Phys., Lpz.*, **53**, 177.
*FÜRTH, R., 1920. *Schwankungserscheinungen in der Physik* (Braunschweig: Sammlung Vieweg).
FÜRTH, R. and MACDONALD, D. K. C., 1946. *Nature, Lond.*, **157**, 807.
JOHNSON, J. B., 1925. *Phys. Rev.*, **26**, 71.
KAPPLER, E., 1931. *Ann. Phys., Lpz.*, **11**, 233; 1932. *Ibid.*, **15**, 545.
MOULLIN, E. B., 1938. *Spontaneous Fluctuations of Voltage* (Oxford University Press).
ORNSTEIN, L. S., 1927. *Z. Phys.*, **41**, 848.
ORNSTEIN, BURGER and TAYLOR, 1927. *Proc. Roy. Soc.*, A, **115**, 391.

* These references deal generally with the subject.

ORNSTEIN, L. S., and MILATZ, J. M. W., 1938. *Physica*, **5**, 971.

RICE, S. O., 1944. *Bell Syst. Techn. J.*, **23**, 282 ; 1945. *Ibid.*, **25**, 46.

*v. SMOLUCHOWSKI, M., 1923. " Abhandlungen über die Brownsche Bewegung und verwandte Erscheinungen " (Leipzig : Ostwald's *Klassiker der exakten Naturwissenschaften*, Nr. 207).

*UHLENBECK, G. E. and ORNSTEIN, L. S., 1930. *Phys. Rev.*, **36**, 823.

*WANG, M. E. and UHLENBECK, G. E., 1945. *Rev. Mod. Phys.*, **17**, 323.

WHITTAKER, E. T. and WATSON, G. N., 1935. *Modern Analysis* (Cambridge : The University Press), p. 384.

* These references deal generally with the subject

DISCUSSION

on the foregoing papers by G. G. MACFARLANE (p. 366) and R. FÜRTH and D. K. C. MACDONALD (pp. 375 and 388).

Mr. D. A. BELL. As the paper by MacDonald and Fürth is concerned with a special case where " shot noise " and " Johnson noise " formulae are easily made to give the same answer, I propose to review the general relationship between these phenomena. They are in fact identical to the extent that both arise from the transport of electric current over a given path by a number of discrete units of charge travelling over limited paths, according to the equivalence

$$l\times i\equiv\sum_{q}\sum_{\Delta x} q\,\frac{\Delta x}{\Delta t} \qquad \ldots.(1)$$

(l=length of path over which current i flows, q=magnitude of an individual charge, Δx=length parallel to l of an individual path, Δt=time which q takes to travel Δx, and the double summation is to cover all charges present in the system and all paths traversed by each charge in unit time). In most cases, the current can be divided into a mean component and a superimposed fluctuation, corresponding to a division of $\Sigma\Delta x/\Delta t$ into a drift velocity and a thermal agitation velocity. Applying this to a metallic conductor, the quantities $\boldsymbol{e}/\boldsymbol{m}$ of the electron, length Δx of the free path and number N of electrons present can be eliminated in terms of the resistance R, which is a measure of the drift component of the double summation on the R.H.S. of (1) ; and assuming an equipartition value $3\boldsymbol{k}\theta/2$ for the total energy of each electron, we find that the mean square fluctuation of current is $\overline{i^2}=(4\boldsymbol{k}\theta/R)df$ (Bell, 1938), in accordance with Nyquists's formula. The proof can also be obtained with any other distribution of electronic energy (Bakker and Heller, 1939). Turning now to the temperature-limited diode, $q=\boldsymbol{e}$, Δx=cathode to anode distance, Δt is vanishingly small, N is known from the mean current (instead of being measured by a resistance), and we obtain Schottky's formula $\overline{i^2}=2i\boldsymbol{e}df$. In the space-charge-limited diode, Δx is still constant ; but Δt, the transit time, is a function both of the anode potential V_{a} and of the initial thermal energy of the electrons, which has a part $\boldsymbol{k}\theta/\boldsymbol{e}$ electron volts corresponding to the component of velocity directed towards the anode. The relation between $\overline{i^2}$ and i in a space-charge-limited valve is then a function of the ratio $\boldsymbol{e}V_{\mathrm{a}}/\boldsymbol{k}\theta$ (Williams, 1941). In the space-charge-limited diode, Δx is the electrode spacing, N is known from the mean current, and Δt can be calculated for plane and for cylindrical diodes (Bell, 1942) : it is the transit time making due allowance for the initial velocities. One can then calculate the fluctuation current corresponding to any given values of mean current and $\boldsymbol{e}V_{\mathrm{a}}\boldsymbol{k}/\theta$, on the assumption that the thermal components of velocity are responsible for the fluctuation, the mean current being maintained constant by the controlling effect of the space-charge and potential barrier. Pierce has pointed out (Pierce: in Bell, 1943) that by regarding as effective only the difference between the individual thermal velocities and the mean forward component of all thermal velocities, my method of calculation is equivalent to the diode having Johnson noise with a temperature equal to 0·644 times cathode temperature, which is in agreement with other theoretical calculations (Rack, 1938). On the other hand, using the full magnitude of thermal velocities in the formula for cylindrical diodes, I get results in reasonable agreement with published measurements of noise in diodes. The retarding-field condition, which is investigated by the authors, is unique in that agreement of the shot noise expression with the Johnson noise expression requires the electron stream to have a temperature exactly half the cathode temperature. This is a case which I have not worked out from first principles ; but it may be said that the only practical way of observing the temperature of the electron stream is by measuring the noise, since it is only through the noise that we can get that interchange of energy between the electron stream and an external body which is necessary for thermometry. With this proviso, there can be no disputing that the noise generated in the current stream of a valve having a finite internal resistance may properly be expressed as a thermal noise conforming to Nyquist's theorem.

As regards Dr. Macfarlane's paper, it was stated that the apparent relaxation time, which appears in the relation between noise amplitude and frequency, corresponded to a periodicity of the order of one cycle per hour. Is it suggested that, in the thermionic flicker effect for example, the diffusion of ions proceeds at this sort of speed ? I am not satisfied that the noise in poor conductors, such as carbon, is likely to be of the same origin as in barrier-layer semi-conductors having rectifying and photoelectric properties which are known to depend on phenomena on an atomic scale. Could not the noise in carbon resistors arise from macroscopic contact variations at the boundaries between granules, not involving the formation and loss of ions ? Again, the linear increase of noise with period of the measuring apparatus seems to me to indicate merely some degree of positive correlation between current density and conductivity : any fluctuation then tends to be self-maintaining, so that long pulses are more probable than short pulses. The mechanism of such a correlation may differ in different substances.

Dr. C. S. **Bull**. The strong resemblance between the equations for the noise of resistances and that of some valves tempts one to speculate on the connection between the two phenomena.

Nyquist's equation is derived, and is valid only, for a resistance of any kind in thermal equilibrium with its surroundings. Such equilibrium permits the random variation of the energy of the resistance by sharing its energy with the much larger amount contained in the constant temperature surroundings. The energy of the resistive circuit itself is therefore not conserved.

In a valve, however, except in one special case, there never exists thermal equilibrium with the surroundings. In fact, it is easy to follow the energy exchanges arising from the passage of electrons across the circuit capacity, the loss of energy from the cathode due to the emission of electrons, and the power expended by the noise voltage in complete detail by applying the principle of conservation of energy to the valve and its circuit. The result of such an investigation produces the well known equations for the full shot noise and for the noise of a valve in the retarding region of the characteristic. The principles applied are so different from those used in deriving Nyquists' equation that the resemblance of some of the resulting equations does nothing to indicate any possible connection between the two processes.

Turning to the special case discussed by D. O. North in his papers on noise, about 1940, we find that he contemplated a valve with two isolated electrodes in a constant temperature enclosure, one electrode (the cathode) having a lower work function than the other (the anode). It is easy to show that an impedance exists between these electrodes, and that the full shot noise of the equal and opposite currents from cathode to anode and anode to cathode generates the noise voltage given by Nyquist's equation for this valve. This result, of course, could have been expected, since Nyquist's theory puts no restriction on the nature of the resistance. At this point, North removes the anode-cathode current, thereby converting the system into an ordinary valve in the retarding region, and obtaining the well known equation for the noise of such a valve. Physically, such a change in the anode current can only be brought about by the destruction of the thermal equilibrium so necessary to the application of Nyquists' equation to the valve. North's special valve therefore fails also to demonstrate any possible connection between the noise fluctuations of valves and circuits.

Mr. D. A. **Wright**. The physical processes envisaged by Dr. Macfarlane as accompanying current flow through the semi-conductor lead to a law relating noise, current density and frequency which has considerable experimental support. I should like to point out, however, that there are some difficulties associated with this picture. It is essential to the argument that clusters of atoms migrate to the anode side of the semi-conductor and there spread over the surface. Before their arrival, electrons were not readily emitted from that part of the sufrace, but following their arrival, and until their loss by diffusion away as ions, the electron emission is enhanced. Now there are several arguments, which we cannot develop here, against the view commonly held that emission from an oxide coating is dependent on barium atoms adsorbed on the outer coating surface. However, in the case of a semi-conductor with direct contacts, the difficulties become quite clear. The theory requires that the flow through the semi-conductor is controlled by the transfer of electrons from semi-conductor to metal at the positive electrode. When adsorbed atoms are present on the surface of the semi-conductor, this transfer is increased. Thus the theory supposes there is no limitation of flow at the negative electrode where electrons pass from metal to semi-conductor. Thus when the system has rectifying properties, the direction of difficult flow for electrons must correspond with electron flow from semi-conductor to metal. This is contrary to the known direction of rectification in the case of an excess semi-conductor such as zinc oxide. This objection would therefore apply equally to barium oxide.

To conform with the experimental results and theories of contacts between metals and semi-conductors it is necessary to suppose that the current flow is determined at the boundary where electrons pass from metal to semi-conductor, and we might expect the noise to be determined similarly by conditions at this boundary. While details of the arrival and loss of ions and atoms at this boundary are probably very important, it is not easy to see how cluster movement could occur, and it is therefore not easy to apply Dr. Macfarlane's theory, at least if we confine our attention to the movements of the metal ions and atoms. It is just possible that oxygen movement might occur in clusters. In the absence of clusters, the theory would predict again a law of the Schottky type.

Mr. Z. Jelonek. When a theoretical formula is checked experimentally, it is important to consider what extent of agreement can be expected. The correlation functions of the noise amplitude derived theoretically and from the experimental record show good agreement. In this connection I should like to enquire, firstly, how and with what accuracy the Fourier transformation necessary to derive the correlation function from the frequency response of the amplifier network was obtained. Secondly, what was the expected accuracy of the derivation of the correlation function from the noise record. It is well known that in order to obtain good accuracy a very great number of readings must be taken. For example, if a random variable has Gaussian probability distribution, and if an accuracy about $\pm 0{\cdot}01$ in the product moment correlation coefficient is required, it may be necessary to take as many as 10,000 readings. (See, for example, *The Advanced Theory of Statistics*, by M. G. Kendall, vol. i, p. 211.)

In view of such consideration I am interested to know whether the procedure used was sufficiently detailed to make the good agreement shown really significant.

Mr. R. Dehn. I should like to ask Dr. Macfarlane if any reason can be given for the *decrease* of the described type of noise with *increasing* temperature.

Mr. B. N. Watts. I should like to ask Dr. Macfarlane if there is a correlation between the value of x in his formula z^{x+1}/f^{x} and the method of deposition, and if it is affected by the presence of the photo-conductivity effect in the film.

Dr. M. Pirenne. I should like to ask Dr. Fürth and Dr. MacDonald whether their experiments give information about statistical fluctuations in the number of light quanta acting upon the photocell. Is the outcome of the experiments the same as if the number of quanta acting upon the photocell, and the number of photo-electrons liberated, were perfectly constant in time?

When variations in the nominal light intensity are produced by the experimenter, these artificial variations of course can have an effect on the electric current of the apparatus. Can the natural fluctuations of intensity—due to the quantum nature of light—alter the characteristics of the electric current, or not? If the case were comparable to that of a Geiger-Müller counter for X-ray photons, these fluctuations should affect the current.

Dr. E. G. James. Dr. Macfarlane has stated that the migration of clusters of atoms to the surface of the cathode coating accounts for the low-frequency noise in thermionic valves, and that the same mechanism would explain the excess noise found in the other semi-conductors. The " flicker " effect associated with oxide cathodes is only noticed in measurements made at audio- or very low radio-frequency, while in certain semi-conductors, e.g. a tungsten-silicon rectifier, excess noise is found to exist at frequencies up to several megacycles per sec. Is it possible to explain this difference on the above theory ?

Another possible mechanism which would give rise to excess noise at low frequency is the movement of ions across the potential barrier between the metal and semi-conductor. These ions would either increase or depress the potential at the barrier, and would cause a fluctuation in the current, the frequency spectrum of which would depend on the mobility of the ions.

As Mr. Wright has mentioned, there is evidence of a potential barrier between the cathode coating and the cathode core in a thermionic valve, and the fluctuation of this barrier may be the source of the low-frequency noise in this case also.

Dr. G. G. MACFARLANE. In reply to Mr. Bell, the figure of one hour for the relaxation time refers to a semi-conductor such as lead sulphide at room temperature. In a thermionic emitter at, say, 950° K. it is of the order of 1/400 sec. (Sproull, 1945). This corresponds to the average time taken to ionize a cluster of impurity atoms which have diffused on to the surface of a contact. It does not by itself tell us anything about the rate of diffusion of ions, since we do not know how far an ion must be from the surface before its dipolar moment ceases to influence the work function.

Regarding Mr. Wright's objection to the mechanism proposed, it is important to differentiate between the total flow of electrons and the noise fluctuations. In the case of an excess semi-conductor with direct contacts to metal it is true that the mean current is largely controlled by the contact where electrons flow from metal to semi-conductor, but fluctuations can still occur due to adatoms at the other contact, provided these alter the work function at the contact. The magnitude of the fluctuations will be smaller than if the clusters of adatoms occurred at the high-resistance contact, but the frequency spectrum will be the same.

In reply to Mr. Dehn, I would like to correct any impression I may have given that the noise increases with temperature without any further restriction. The statement is only true in the theory provided the mean current is kept constant. This behaviour is expected for the following reasons :—

(i) In the formula for $\overline{\Delta j_2}$ the factors most sensitive to temperature are q^{2m} and $1/\omega^{2m}$.

(ii) The index m is a function of a, which, being Langmuir's constant, is proportional to $1/T$. Therefore since m increases as a is reduced, it also increases as T is increased.

(iii) The decay constant q depends on T through a factor of the form $\exp(-T_0/T)$, as shown on page 373. Therefore we can write

$$\begin{aligned}\overline{\Delta j_2}/j_1 &\propto q^{2m}/\omega^{2m}\\ &\propto \exp\{-2m(T_0/T+\log\omega)\}.\end{aligned}$$

In view of the increase of m with T, the index $2m(T_0/T+\log\omega)$is found to be a function that increases with T, so that the normalized power density of the noise decreases with rise of temperature.

In reply to Mr. Watts, I do not know of any definite correlation between the index x and either the method of preparation or the photo-conductivity.

I am not sure whether the excess noise in tungsten-silicon rectifiers mentioned by Mr. James shows the type of frequency spectrum associated with flicker noise. Noise in excess of Johnson noise does occur at these frequencies in most semi-conductors, due to fluctuations in the number of electrons in the conduction band, but its spectral power density does not fall off rapidly with increasing frequency.

Dr. R. FÜRTH. Several attempts have been made in the past to prove the existence of fluctuations in radiation by measuring fluctuations in the emission current of photoelectric cells, as suggested by Dr. Pirenne. But the objection to such experiments is that a photocell must exhibit fluctuations even in the complete absence of radiation fluctuations, owing to the fact that the relation between the incoming photons and the emitted electrons is of a statistical nature. More recently one of my collaborators (A. Kolin, *Ann. Phys., Lpz.*, **21**, 813 (1933)) used an ultra-violet counter for a similar experiment (i.e. a Geiger-Müller counter with an ultra-violet sensitive cylinder) which permits one to count ultra-violet photons separately. However, it appeared from a statistical analysis of the time-interval distributions of the counts that here too the effect of the fluctuations in the number of incoming photons is completely masked by the efficiency fluctuations of the counter. We may therefore conclude that the effect of statistical fluctuations of the illuminating light on the experiments of Macdonald can be disregarded.

Dr. D. K. C. MACDONALD. In reply to Mr. Jelonek, the Fourier transformation involved in the Wiener-Khintchine theorem was obtained on the assumption that the overall frequency-response of the amplifier network was Gaussian. This is a customary assumption for a network of relatively narrow band-width composed of a number of tuned circuits in cascade, which agrees well in practice. The Gaussian parameter was determined by numerical integration of the response curves ; it was felt that this method (rather than, for example, logarithmic plotting of the response curve) would best give effect to the whole response curve.

As regards the second point, it should be emphasized that in our work a *difference* correlation-coefficient was employed throughout. Consequently, for the larger values of correlation-interval, where the examination of the agreement between theory and experiment becomes particularly interesting, the absolute magnitude of the correlation coefficient is relatively large (tending to a constant value for infinite interval) ; therefore the proportionate accuracy obtainable is quite adequate with a reasonable number of observations. We may further mention that 200–300 primary observations were used on each record, and in all some eight records were examined in statistical detail ; we were entirely satisfied that the good agreement obtained overall and in individual records was significant.

In confirming Dr. Fürth's remarks relative to Dr. Pirenne's question, I should also like to draw attention to a similar problem relating to the voltage fluctuations of a piezo-electric crystal discussed recently elsewhere (Lawson and Long, 1946 ; Brown and MacDonald, 1946). In that case the relationship of the electrical fluctuations to the concomitant spontaneous mechanical vibrations was under review ; it was there emphasized that in thermal equilibrium the magnitude of the electrical fluctuations generated is completely determined by the electrical characteristics of the system (Nyquist's theorem) independent of the particular electro-mechanical interaction involved.

As regards the points raised by Mr. D. A. Bell and Dr. C. S. Ball, it is, in my opinion, evident that in general Nyquist's theorem cannot be applied directly to a thermionic valve as a whole to determine the noise generated at the external terminals of the valve. It does, however, appear to me entirely reasonable to ascribe to the cathode-potential-barrier region ***generally*** a ***short-circuit*** noise current given by

$$\overline{(i-I)_b^2}=\tfrac{1}{2}(4g_b\boldsymbol{k}T\Delta f), \qquad \ldots\ldots(1)$$

where g_b is the barrier-cathode differential conductance. On applying Helmholtz's (Thévenin's) constant-current theorem, the short-circuit noise current obtaining at the external (cathode-anode) terminals will be less than (1) by a formally simple impedance transformation factor involving g_b and the barrier-anode impedance. Noting that in the parallel-plane structure $g_b=\dfrac{eI}{\boldsymbol{k}T}$, it follows that the " space-charge reduction factor ", Γ^2, is fundamentally no more than this impedance transformation. Schottky (1936) formulated the impedance-transformation concept in a somewhat lengthy discussion, but was in error in evaluating it. It is clear (MacDonald, 1946) that when the barrier : anode impedance is evaluated on the basis of the Maxwell-Boltzmann law, this view-point leads to precisely the same result as that obtained by North (1940), Rack (1938) and Schottky in his later work (1937). The actual numerical computation (as undertaken by these workers) is naturally rather complex, but the same is true of the evaluation of any other theoretical valve parameter when account is taken of the velocity emission law, and it should not be inferred that the noise problem itself presents any greater difficulty in principle. In this sense, and to this extent then, I believe that " shot " and " thermal " noise may be unified.

REFERENCES

Bakker, C. J. and Heller, G., 1939. *Physica*, **6**, 262.
Bell, D. A., 1938. *J. Instn. Elect. Engrs.*, **82**, 522.
Bell, D. A., 1942. *J. Instn. Elect. Engrs.*, **89**, 207.
Brown, J. B. and MacDonald, D. K. C., 1946. *Phys. Rev.*, **70**, 976.
Lawson, A. W. and Long, E. A., 1946. *Phys. Rev.*, **70**, 220, 977.
MacDonald, D. K. C., 1941. Ph.D. Thesis, Cap. 2, pp. 31–2 (University of Edinburgh).
North, D. O. *et al.*, 1940. *R.A.C. Review*, **4**.
Pierce, J. R., 1943. (Discussion on Bell, 1942.) *J. Instn. Elect. Engrs.*, **90**, 148.
Rack, A. J., 1938. *Bell Syst. Tech. J.*, **17**, 592.
Schottky, W. *et al.*, 1937. *Wiss. Veröff. Siemens-Werk.*, **16**, 1.
Schottky, W., 1938. *Die Tel. Röhre*, **8**, 175.
Williams, F. C., 1941. *J. Instn. Elect. Engrs.*, **88**, 219.

THE MASS OF THE NEUTRINO

By F. C. FRANK,
H. H. Wills Physical Laboratory, Bristol

MS. received 6 *February* 1946; *in revised form* 16 *December* 1946

ABSTRACT. From a collation of the energies involved in nuclear reactions and radioactive decays, which together represent closed cycles, it is shown that the rest-mass of the neutrino, m_ν, is beyond any reasonable doubt less than that of the electron, m_e, and probably less than $m_e/10$. It cannot be shown with certainty to be measurably distinct from zero.

§ 1. INTRODUCTION

In Mattauch and Flügge's *Kernphysikalische Tabellen* (1942) there is only one statement concerning the rest-mass of the neutrino, viz. that on p. 76 that it is certainly not larger, and is probably smaller, than that of an electron. This, however, is an almost tendentiously conservative statement: it is clearly a question of considerable significance whether the mass of the neutrino, m_ν, is of similar magnitude to that of the electron, m_e, or no. Such experimental evidence as we possess, which is presented elsewhere in Mattauch and Flügge's Tables, indicates fairly unambiguously that it is not. This has been the usual view adopted in theoretical considerations of the matter. F. Perrin (1933) surmised on qualitative grounds that the rest-mass of the neutrino ought to be zero to account for the average value of the β-energy. Bethe and Bacher (1940) deduced from the form of the β-spectrum, in connection with Fermi's theory (1934), that the neutrino mass is not more than one-fifth that of the electron. Bethe and Bacher also reached a similar conclusion from another argument, based on the number of existing pairs of adjacent odd isobars (i.e. nuclei of the same odd mass number, differing in atomic number by 1). It is observed that the energies of odd isobars, plotted against atomic number, lie close to smooth parabolae, the form of which depends on mass number in a known manner. Thus it is possible to make a statistical estimate of the chance that for some odd mass number in the range of the natural nuclei there exists a pair of odd isobars, being the two most stable nuclei for their mass number, differing in energy by less than any stated amount.

It also appears certain that any nucleus will sooner or later undergo β-decay (positive or negative) or will capture one of its planetary electrons, if energy can be released in the process. It is postulated that each such process involves the emission of a neutrino. Then the two adjacent isobars will both be stable if their energies differ by less than the energy equivalent of m_ν. If $m_\nu=0$, there will be no stability range at all between the reverse processes of electron emission and electron capture, and therefore no stable pairs of adjacent isobars. In fact, there are three pairs of adjacent odd isobars not known to undergo change, viz. Cd—In, mass number 113; In—Sn, 115; and Sb—Te, 123. A fourth pair, Re—Os, 187, listed by Mattauch and Flügge, has since been eliminated (Lougher and Rowlands, 1944). If it is assumed that these three pairs are genuinely stable, for the reason mentioned, we have the statistical estimate $m_\nu c^2 \approx (0{\cdot}03 \pm 0{\cdot}02)$ Mev. If m_ν were as large as m_e, we should expect about 50 such pairs. Thus this argument definitely indicates that the mass of the neutrino is much less than that of the electron. On the other hand a similar calculation shows that there is a statistical likelihood of about two pairs differing in energy by less than the binding energy of a K-electron ($13{\cdot}6\,Z^2$ electron volts). For such nuclei, only electrons from remoter shells are available for capture, and the capture-decay is bound to be slow and difficult to detect. Flügge, in the tables mentioned, summarizes reasons, based on isotopic abundances, for supposing that such undetected changes may occur in these cases. Hence this argument does not exclude the value $m_\nu=0$.

All of the above arguments are somewhat oblique, and this paper is primarily concerned with more nearly direct estimates of m_ν, from the energies of nuclear processes, using the data available in Mattauch and Flügge's Tables.

§ 2. CYCLE ENERGIES

The mass of the neutrino can be found most directly from the energy balance of simple cycles of nuclear reaction followed by β-radioactive decay, i.e., by the generation of a positive or negative electron, and simultaneously, according to the hypothesis, of a neutrino. There may also, sometimes, be γ-emission to complete the cycle. The simplest types of cycle are accordingly:

$$\text{(I)}\quad {}^{y}_{x}At\,(n,p)\,{}^{y}_{x-1}At'\,(-,e^-\nu\gamma)\,{}^{y}_{x}At,$$

$$\text{(II)}\quad {}^{y}_{x}At\,(p,n)\,{}^{y}_{x+1}At'\,(-,e^+\nu\gamma)\,{}^{y}_{x}At.$$

The notation employed here corresponds to that of Mattauch and Flügge. ${}^{y}_{x}At$ signifies the atomic species At, of atomic number $Z=x$ and mass number $A=y$. Brackets signify a nuclear reaction, wherein what appears before the comma enters the nucleus, and what appears after the comma emerges from it. A hyphen before the comma signifies a spontaneous reaction—i.e. a radioactive decay.

If $Q_{(n,p)}$ or $Q_{(p,n)}$ is the energy yield of the corresponding nuclear reaction stage of the cycle, Q_β is the limiting energy of the β rays and Q_γ the energy of the γ rays emitted, if any, the mass of the neutrino is to be calculated as

$$m_\nu = m_n - m_p - m_e - \frac{1}{c^2}\,(Q_{(n,p)} + Q_{\beta^-} + Q_\gamma), \qquad \ldots\ldots(1)$$

$$m_\nu = m_p - m_n - m_e - \frac{1}{c}\,(Q_{(p,n)} + Q_{\beta^+} + Q_\gamma). \qquad \ldots\ldots(2)$$

There are additional cases in which, although the nuclear reaction (n, p) or (p, n) is not realized directly, its energy yield can be determined from the two reactions

$${}^{y-1}_{x}At''(d, p)\,{}^{y}_{x}At \quad \text{and} \quad {}^{y-1}_{x}At''(d, n)\,{}^{y}_{x+1}At',$$

i.e. reactions in which bombardment with deuterons causes emission of protons and neutrons respectively. Cycles thus realized indirectly can be called types (I *a*) and (II *a*).

Cycles which are completed by electron capture instead of radioactive emission (type III):

$${}^{y}_{x}At\,(p, n)\,{}^{y}_{x+1}\,At'\,(e_{K,}\,\nu\gamma)\,{}^{y}_{x}At,$$

only allow us to find an upper limit to m_ν, since there is no means of measuring the energy carried away by the neutrino,

$$m_\nu \leqslant m_p - m_n + m_e - \frac{1}{c^2}(Q_{(p,\,n)} + Q_\gamma + Q_x). \qquad \ldots\ldots(3)$$

where Q_x is the energy appearing as X rays when external electrons fill the place of the K-electrons absorbed.

The difference $(m_n - m_p)$ could in principle be eliminated by combining equations (1), and (2), but this is unnecessary and wasteful of measurements, as it is known to a higher accuracy than cycle energies, from mass spectroscopy combined with measurement of the energy involved in the reaction $D(\gamma, n)H$. From Mattauch's best selected data it is

$$m_n - m_p = 1{\cdot}361 \pm 0{\cdot}025 \text{ TMU}.$$

(Mass is quoted throughout this note in TMU—thousandths of a mass unit, called TME by Mattauch and Flügge; 1 TMU is 1/16,000 of the mass of a neutral atom of ${}^{16}_{8}O$, and its equivalent energy according to $E = mc^2$ is 0·931 Mev.) The mass of the electron, m_e, is 0·547 TMU, and should be free from error to this number of significant figures. In terms of energy,

$$\begin{aligned} m_e c^2 &= 0{\cdot}509 \text{ Mev.}, \\ (m_n - m_p)\,c^2 &= 1{\cdot}267 \pm 0{\cdot}024 \text{ Mev.}, \\ (m_n - m_p - m_e)\,c^2 &= 0{\cdot}758 \pm 0{\cdot}024 \text{ Mev.}, \\ (m_p - m_n - m_e)\,c^2 &= -1{\cdot}776 \pm 0{\cdot}024 \text{ Mev.} \end{aligned}$$

Table 1 summarizes data for all cycles of the specified types for which there is information in Mattauch and Flügge, Tables IV and VI. Table IV contains, *inter alia*, the limiting energies of β-spectra, which are here entered in column 3. In one case (the β-decay of ${}^{20}_{9}F$) the energy of a γ ray has been added in as well. Table VI collects together measurements of the energy yields of nuclear reactions, which are here entered in column 2. These include some values derived by multiplying the threshold energy for the reaction (p, n) by $A/(A+1)$. In both columns, single weighted mean values are entered in the few cases in which the tables list two or more measurements of comparable precision: in such cases no weight is given to measurements with no declared probable error. The remaining two columns of table 1 are calculated from columns 2 and 3.

Table 1

Nuclei involved in cycle	Q reaction (Mev.)	Q decay (Mev.)	Q cycle (Mev.)	$m_\nu c^2$ (Mev.)
Type I : $(n, p)-\beta^-$				
${}^{14}_{7}\mathrm{N}$ ${}^{14}_{6}\mathrm{C}$	0·55±0·03	0·12±0·02	0·67±0·037	0·088±0·05
Type I *a* : the same from $(d, p)-(d, n)$				
${}^{3}_{2}\mathrm{He}$ ${}^{3}_{1}\mathrm{H}$	0·67±0·036	0·01±0·002	0·68±0·036	0·078±0·043
(from ${}^{2}_{1}\mathrm{D}$) (3·98±0·02−3·31±0·03)				
${}^{10}_{5}\mathrm{B}$ ${}^{10}_{4}\mathrm{Be}$	0·35	0·55	0·90	−0·142
(from ${}^{9}_{4}\mathrm{Be}$) (4·55−4·20)				
${}^{20}_{10}\mathrm{Ne}$ ${}^{20}_{9}\mathrm{F}$.	−6·5	7·2	0·7	0·058
(from ${}^{19}_{9}\mathrm{F}$) (4·3−10·8±0·2) (5·0±2·2)				
Type II : $(p, n)-\beta^+$				
${}^{10}_{5}\mathrm{B}$ ${}^{10}_{6}\mathrm{C}$	−5·1	3·36±0·1	−1·74	−0·036
${}^{11}_{5}\mathrm{B}$ ${}^{11}_{6}\mathrm{C}$	−2·72±0·01	0·981±0·005	−1·739±0·011	−0·037±0·03
${}^{13}_{6}\mathrm{C}$ ${}^{13}_{7}\mathrm{N}$	−2·97±0·03	1·21±0·01	−1·76±0·032	−0·018±0·04
${}^{18}_{8}\mathrm{O}$ ${}^{18}_{9}\mathrm{F}$	−2·42±0·04	0·7	−1·72	−0·056
${}^{19}_{9}\mathrm{F}$ ${}^{19}_{10}\mathrm{Ne}$	−3·97	2·20	−1·77	−0·006
${}^{23}_{11}\mathrm{Na}$ ${}^{23}_{12}\mathrm{Mg}$	−4·58±0·3	2·82	−1·76	−0·016
${}^{27}_{13}\mathrm{Al}$ ${}^{27}_{14}\mathrm{Si}$	−5·8±0·1	3·64±0·1	−2·16±0·14	0·384±0·15
${}^{61}_{28}\mathrm{Ni}$ ${}^{61}_{29}\mathrm{Cu}$	−3·0	0·94	−2·06	0·284
${}^{64}_{28}\mathrm{Ni}$ ${}^{64}_{29}\mathrm{Cu}$	−2·5	0·655±0·003	−1·845	0·069
${}^{63}_{29}\mathrm{Cu}$ ${}^{63}_{30}\mathrm{Zn}$	−4·0	2·32±0·005	−1·68	−0·096
${}^{68}_{30}\mathrm{Zn}$ ${}^{68}_{31}\mathrm{Ga}$	−3·6	1·85	−1·75	−0·026
Type II *a* : the same from $(d, n)-(d, p)$				
${}^{11}_{5}\mathrm{B}$ ${}^{11}_{6}\mathrm{C}$	−3·06	0·981±0·005	−2·079	0·303
(from ${}^{10}_{5}\mathrm{B}$) (6·08−9·14±0·06)				
${}^{13}_{6}\mathrm{C}$ ${}^{13}_{7}\mathrm{N}$	−2·96±0·06	1·212±0·004	−1·748±0·06	−0·028±0·065
(from ${}^{12}_{6}\mathrm{C}$) (−0·25±0·03−2·71±0·05)				
${}^{17}_{8}\mathrm{O}$ ${}^{17}_{9}\mathrm{F}$	−3·65	2·1	−1·55	−0·226
(from ${}^{16}_{8}\mathrm{O}$) (−1·7−1·95±0·06)				
Type III : $(p, n)-K$ (giving upper limit only)				
${}^{7}_{3}\mathrm{Li}$ ${}^{7}_{4}\mathrm{Be}$	−1·62±0·02	0·425±0·025	−1·195±0·032	⩽0·437±0·04

Consistency data ("probable errors") have been included in the table where they are listed, and carried through to the final column, but they have not been utilized for the purpose of weighting the results. Various authors adopt various conventions about the inclusion in the "probable error" of uncertainty in basic calibrations: in general they are a good index of the scatter in measurements, but a poor index of the possible systematic error. Accordingly the 18 cases of types I, I *a*, II and II *a*, have been treated as a homogeneous set of measurements, with a Gaussian expectation of errors. The probable error for each cycle is thus found to be 0·105 Mev., and the mean result with probable error (incorporating the probable error of $m_n - m_p$)

$$\begin{aligned} m_\nu c^2 &= 0{\cdot}0321 \pm 0{\cdot}0357 \text{ Mev.,} \\ m_\nu &= 0{\cdot}0345 \pm 0{\cdot}0383 \text{ TMU} \\ &= (0{\cdot}0631 \pm 0{\cdot}0701)\, m_e. \end{aligned}$$

This statement of "probable error" is subject to the same proviso as that made above—it assumes that errors are unbiased. Two probable sources of bias—errors in locating the limit of a β-ray continuum, and failure to detect a soft γ ray—are both liable to lead to an over-estimate, rather than an under-estimate, of the neutrino mass: a third, over-estimate of the threshold energy for (p, n), would have the opposite effect. This latter error must be larger than all estimated probable errors if the mass of the neutrino is of similar magnitude to that of the electron. In 19 different cycles there is not one which indicates that m_ν is as large as m_e, and m_e is outside the estimated probable error of m_ν by a large factor. Flügge's statement in the tables by Mattauch and Flügge, p. 76, must therefore be rejected as unreasonably conservative. Considering the other evidence as well, the rest-mass of the neutrino is almost certainly less than one-tenth that of the electron, and, so far as present measurements can show, not appreciably different from zero.

REFERENCES

Bethe and Bacher, 1940. *Rev. Mod. Phys.*, **8,** 189.
Fermi, 1934. *Z. Phys.*, **88,** 161.
Lougher and Rowlands, 1944. *Nature, Lond.*, **153,** 374.
Mattauch and Flügge, 1942. *Kernphysikalische Tabellen* (Berlin : Springer).
Perrin, F., 1933. *C.R. Acad. Sci., Paris*, **197,** 1625.

WHAT EXPERIMENTS ARE NEEDED IN FUNDAMENTAL PHYSICS ?

By R. E. PEIERLS,
University of Birmingham

Read 20 *December* 1946; *MS. received* 10 *January* 1947

ABSTRACT. A brief outline is given of the chief outstanding difficulties of fundamental theory concerning the nature, origin and properties of nuclear forces, the properties of mesons, their inter-relation with other particles and their connection with nuclear forces, the lifetime of the meson and its relation to ordinary beta-processes, with the classification of atomic energy levels. A discussion is given of some experiments which might become possible with modern technique and which are likely to help provide the answer.

§1. GENERAL

In discussing the value of possible experiments in fundamental physics from the point of view of present theory, it is not my intention to suggest that it is always possible to predict the kind of experiment that will lead to an important advance in our knowledge. Indeed, from past experience, we must certainly be prepared for further unsuspected discoveries which can only come from a broad study of fundamental phenomena over a wide field. As a typical example of this kind of discovery I need only quote the discovery of x rays; Roentgen certainly set out to study in a general way what happens in a cathode-ray tube, but what precisely he was going to find he did not know in advance. On the other hand we have discoveries like Rutherford's law of the scattering of alpha particles; this was the result of a deliberate attempt to explore the inside of atoms by that particular tool.

It is only experiments of this latter type that are amenable to discussion from a theoretical point of view, but this should not be understood as an attempt to minimize the importance of new and unpredicted observations.

§2. NUCLEAR FORCES

Progress in the theory of nuclei is hampered by lack of knowledge of the precise nature of the forces acting between elementary particles. In this respect the situation differs from that of atomic theory, where, as soon as the existence of the electron and the nucleus was discovered, one could take it for granted that the forces between them were electric in nature and were essentially given by Coulomb's law. The difficulty was there in the change of the general laws of mechanics that found its expression in quantum theory.

In the case of the nucleus it is likely that the general laws of quantum mechanics can, to a large extent, be applied to the phenomena inside the nucleus, although there are some doubts on that score, but the nature of the forces is still essentially unknown. The most direct information about the forces must come from the study of two-body problems, in the same way in which, in atomic theory, the hydrogen atom provided the crucial confirmation for Bohr's theory; but while the hydrogen spectrum contains a large amount of simple data, such as the line frequencies, selection rules, Stark and Zeeman effects, etc., the corresponding nuclear two-body problem, i.e. the deuteron, has no excited states and no spectral lines and cannot be influenced to a measurable extent by external fields. We must therefore fall back on the properties of reactions with two bodies, such as the scattering and capture of neutrons by protons and the scattering of protons by protons. It turns out that, given the binding energy of the deuteron and the capture cross-section for thermal neutrons in hydrogen, the behaviour of all collisions of two particles at energies up to a few million volts can be predicted quite well without further assumptions about the details of the law of force. Owing to the fact that these forces act only over very short range, new information can be obtained only from cases in which the de Broglie wave-length is comparable to the range of the forces, and that means at energies of well above 10 Mev. At energies of this order, experiments by Amaldi and his collaborators have tended to show an asymmetry in the scattering which, if confirmed, could be used to decide whether the forces are "ordinary" or "exchange" forces, but the asymmetries in question are very small, and experiments by Powell in Bristol, using a different method, contradict Amaldi's results. Measuring the intensities of neutron beams precisely is always difficult, and more conclusive evidence could be obtained by going to much higher energies, where the expected asymmetries are larger.

In the case of proton-proton scattering, precision measurements are easier and, while the interpretation of the experiments is more difficult because of the presence of the long-range electrical repulsion, in addition to the nuclear forces, the theory of this problem is sufficiently well known to draw definite conclusions as soon as data at energies of the order of 10 Mev. are available. Proton-proton scattering, however, cannot give all the necessary information, since, for example, the distinction between "exchange" and "ordinary" forces does not arise in the case of identical particles.

Experiments at still higher energies, which will be capable of giving information on the forces at extremely short distances, are also of importance since we now know that the law of forces between two nuclear particles contains a directional force rather like the interaction between two electric or magnetic dipoles. The most elementary theory of this directional or "tensor" force leads to a potential varying with the inverse cube of the distance, as in the case of dipoles, and it is known that the singularity at zero distance is then so strong as to make any stationary state of finite binding energy impossible. The law must therefore be more complicated at small radii, and this is just the point where theory meets one of its greatest difficulties.

A different type of problem is associated with those phenomena in which one observes the interaction of protons or neutrons with an external electromagnetic field, such as the capture of neutrons by protons or the photo-dissociation of the deuteron. The particular interest of these problems lies in the fact that, according to the meson theory of the forces, the proton is not just a point charge but spends a fraction of its time as a neutron with a charge spread over a small region in a surrounding field in which it is trying to generate positive mesons. This view is supported by the fact that the magnetic moment of the proton is not just one nuclear magneton (Bohr magneton divided by the proton-electron mass ratio), as Dirac's theory leads one to expect, but considerably larger. This view of the structure of the proton could be confirmed by comparing quantitative results about the emission and absorption of electromagnetic radiation by protons with the theory assuming a point charge.

§ 3. MANY-BODY PROBLEMS

In the problems of heavier nuclei the success of the general picture first proposed by Bohr has shown that a great many features can be qualitatively understood without specific reference to the nature of the forces. It is therefore difficult to use experiments on heavy nuclei for conclusions about the forces. In principle, of course, a quantitative knowledge of the energy levels and other properties of nuclei could be used to test, or even to derive, the quantitative details of the interaction. However, in practice this would be about as difficult as the job of deriving atomic theory from a knowledge of the frequencies and intensities of the lines in the iron spectrum. In a way the problem is even more difficult because, while in the atom the presence of a strong centre of force and the shell structure allow us to treat each electron to a reasonable first approximation as moving in a given field of force, this simplification is of no help in the case of the nucleus. What would, however, be of great assistance is statistical material about the energy levels, symmetry properties, transition probabilities, etc., of the excited levels of very many nuclei. For instance, the fact that a particular nucleus has a large capture cross-section for slow neutrons has very little direct significance, but knowledge of the number of isotopes in a certain part of the periodic system which have cross-sections above a given value can be used to estimate the distribution of excited states of nuclei of that type. In this connection it would be of particular importance to have some studies of nuclei about which as much information as possible is available simultaneously, such as the energy of the level, its angular momentum and parity, the intensity of any gamma-rays and beta-transitions involving this level, information on the internal conversion of the gamma-rays, etc.

In this connection I would like to draw particular attention to the study of excited levels by means of collisions with charged particles (the nuclear analogue of the Franck-Hertz experiments), which needs careful measuring technique but does not require very high energies and which, while unspectacular, would provide most valuable information.

§ 4. BETA-DECAY

Our knowledge of the beta-ray spectrum has improved a great deal and it looks at present as if the simplest form of Fermi's theory accounts for it reasonably well. Even there the form of the theory is not uniquely fixed by the spectrum, but several alternative laws are possible which give the same spectrum but different selection rules and possibly different spectra for "allowed" (i.e. intense) and "forbidden" (i.e. weak) transitions. What is needed here is, therefore, a close study, in suitable cases, of the symmetry type and spin of the initial and final states of the decaying nucleus, and a knowledge of any gamma-rays emitted before or after the beta-ray. The study of the beta-ray spectrum at very low energies seems to be particularly difficult from the point of view of experimental technique, and it would be important to know how far the remaining discrepancies between theory and experiment at the low energy end can be reduced by an improvement in technique.

The important feature of beta-ray theory is the need for assuming a neutrino, and further experiments of the recoil type would put our belief in its existence on a firmer basis and at the same time, by measuring the angular distribution of the neutrinos in relation to the electron, would help to differentiate between possible alternative theories.

I need hardly say that we will never be quite satisfied about the neutrino unless we have succeeded in finding it not only emitted but also absorbed, but for this something entirely new would be required, and this belongs to the type of discovery for which theory can be of little assistance.

§ 5. MESONS

A great advance in our knowledge of fundamental particles can be expected when means become available to produce in the laboratory particles of sufficient energy to generate mesons. Cosmic-ray evidence suggests a meson mass of somewhat over 100 Mev., but it is not clear at present whether all mesons have the same mass or not, and for a full exploration it would be important to have some excess energy in hand. In principle, mesons could be generated as soon as the kinetic energy of the particle exceeds the rest energy of the meson, but, as in the case of pair creation, one would expect that the effective cross-section for this process would be small immediately above the threshold and that reasonable intensities could only be obtained with energies well above the minimum.

As soon as mesons can be made in the laboratory, one will clearly want to study all their properties, including the results of their collisions with nuclei and other particles and processes in which they are absorbed.

At the same time it is important to find out whether, in addition to charged mesons, there exist neutral ones. If the general ideas of the current meson theory of nuclear forces are justified, neutral mesons must exist and must be produced fairly readily in collisions of fast neutrons or protons with each other or with nuclei.

This group of problems is probably the one in which we may expect the most spectacular advances in our fundamental knowledge as soon as machines for generating the necessary energies come into operation.

§6. ELECTRONS AND ELECTROMAGNETIC RADIATION

The problems of the properties of electrons and their interaction with the radiation field at very high energies and for collisions involving very close approach are probably the most important ones in connection with the deadlock in present theory. No adequate theory is available for describing such interactions at all and, while certain ways of calculation are available, which in all likelihood give the right answers for such things as the scattering of fast gamma-rays by electrons, the emission of light and the creation of pairs in electronic collisions, etc., they all contain arbitrary procedures. If any of the present theories were carried out with logical consistency, one would in most cases get either zero or infinity as an answer. It is not easy to suggest in detail what experiments are likely to help in finding a solution of these difficulties, except that the more our knowledge of the phenomena involving fast electrons, the greater the chance of bringing order into the present complex situation.

DISCUSSION

Mr. E. S. SHIRE. May I ask Prof. Peierls at what energies he would expect the current theory of collisions between elementary particles, particularly electrons, to break down ?

Prof. M. L. OLIPHANT. I should like to ask Prof. Peierls whether any phenomena of the nature of " polarization " phenomena are likely to appear with fundamental particles at close distances of approach ?

Dr. O. FRISCH. I agree with Prof. Peierls that much experimental information of a statistical character (level density, distribution of width values etc.) should be collected to increase our knowledge of nuclei. At the same time I want to point out that there are some known experimental facts which indicate the presence of definite structural features even in heavy nuclei. In particular, there are several facts which show that various nuclear properties are very little affected by the precise number of neutrons in the nucleus. The most striking fact of this kind is the strict parallelism of the three well-known natural radioactive series ; they all show alpha decay of increasing violence down to the appropriate lead isotope, when alpha decay suddenly stops, they all show branching at the bismuth isotope, and so on.

Another structural feature shows up if one plots the magnetic moments of nuclei against their angular momenta. Nuclei containing an odd neutron have small magnetic moments, of the order of that of the free neutron, while nuclei containing an odd proton have magnetic moments increasing with the angular momentum. It looks as if the odd nucleon was actually running on an orbit.

Phenomena of that kind raise the hope that a partly structural (and not merely statistical) model of atomic nuclei will eventually emerge.

Prof. P. B. MOON. I would like to call attention to another well known fact that may point to a similarity of structure between nuclei that differ by one or a few neutrons; namely the occurrence of metastable states in several isotopes of a few elements, such as Te and In.

AUTHOR'S REPLY. With regard to Mr. Shire's question, it is not easy to judge what energies should be regarded as " high " for this purpose. One would guess that the order of magnitude should be that energy at which the de Broglie wave-length becomes comparable with the classical electron radius. This is of the order of 137 electron masses. One must, however, remember that we shall get no new information from processes involving fast particles which, after a suitable Lorentz transformation, appear to depend only on phenomena involving slow particles. It is probably right to say that it is not sufficient to have particles of energies of that order, but also to observe with them phenomena in which a close collision of the particles to distances of the order of 10^{-13} cm. is essentially involved. Such processes will, in general, involve cross-sections of the order of 10^{-26} cm^2 or less, and we can expect interesting information only from conditions in which such comparatively rare events are made observable.

In reply to Prof. Oliphant, while some of the current theories talk of a finite size for elementary particles, I would not like to interpret this as a definite structure, since then relativity would make it impossible to assume the structure as rigid, and one would have to admit internal degrees of freedom of such a structure which, upon quantization, would again lead to new particles. Instead, one likes to regard the structure as connected with a limitation in our fundamental concepts about space which might make it impossible to locate the charge belonging to the particle at a precise mathematical point. In this sense the structure is not liable to change under external forces and the question of polarization does not arise.

All current theories, however, involve the assumption that since the electron can virtually generate light quanta and light quanta can generate pairs of electrons, whereas protons and neutrons can generate mesons, all these virtual processes express themselves in disturbances in the space surrounding each particle. I have referred to one example of this in the case of the charge distribution near a proton. These disturbances may themselves be influenced by the presence of the particles and in that sense there may exist polarization effects. However, just on this point the present theory is on rather uncertain ground and it is one of its unsatisfactory features that a phenomenon apparently so simple as a single electron or light quantum or meson, or even empty space, appears in the theory as a state of affairs of enormous complexity.

The kind of consideration to which Dr. Frisch and Professor Moon have referred revealed general regularities in the properties of nuclei, which, I agree, may very well help in forming a more systematic picture of nuclear properties. Unfortunately no-one has, as yet, succeeded in drawing clear conclusions from them. It is true they are compatible with the idea of alpha particles existing as separate sub-units in the nucleus, but they are equally compatible with the opposite extreme of individual neutrons and protons moving in independent orbits, and neither of these extremes is likely to be near the truth.

COLLISION BROADENING OF THE INVERSION SPECTRUM OF AMMONIA AT CENTIMETRE WAVE-LENGTHS. I.—SELF-BROADENING AT HIGH PRESSURE

By B. BLEANEY and R. P. PENROSE,
Clarendon Laboratory, Oxford

MS. received 3 January 1947

ABSTRACT. The absorption spectrum of ammonia between 0·6 and 0·9 cm.$^{-1}$ has been measured at pressures up to 60 cm. Hg. From the previous analysis of the fine structure, and the measurement of the widths of the lines at a pressure of 0·5 mm. Hg, the shape of the absorption curve at high pressures has been computed assuming that the widths vary accurately as the first power of the pressure. At a pressure of 10 cm. Hg, the computed and observed curves agree within the experimental error ; at 60 cm. Hg, the observed attenuation at the lower frequencies is somewhat greater than that computed. Possible reasons for this phenomenon are discussed.

§ 1. INTRODUCTION

The phenomenon of the broadening of spectral lines in an absorbing or emitting gas owing to collisions with other molecules has been studied both in the optical region and in the infra-red. In these regions the contribution to the width of a line due to collision broadening is, at ordinary pressures, of the same order as that due to the Doppler effect. In the region of centimetre wave-lengths, however, the situation is very different. At atmospheric pressure the collision frequency of the molecules of a gas is of the order of 10^{10}, which is comparable with the frequency of the radiation. Spectral lines in this region will therefore be extremely broad at atmospheric pressure. Other contributions to the widths are, on the contrary, very small; the most important is that due to the Doppler effect, which produces a width which is proportional to the frequency of the radiation, and about 10^{-6} times smaller. At pressures down to a few hundredths of a millimetre, therefore, the width of a spectral line at these wave-lengths will be determined solely by collision broadening, and the opportunity arises for studying pressure broadening in some detail. Furthermore, in the centimetre wave-length region, accurate measurements of absorption intensities can be made; thus the shape of a spectral line, and the variation of its width, can be studied over a range of pressures from atmospheric, or greater, down to 0·01 mm. or less. Such pressures have the additional advantage that effects due to multiple encounters should be small, and the complications which these introduce into work at pressures of several atmospheres should therefore be absent.

The only strong resonant absorption at present known to exist at centimetre wave-lengths is that due to the inversion of the ammonia molecule in its ground state. This spectrum, which possesses a considerable fine structure, has been analysed (Bleaney and Penrose, 1946 a, b), using pressures of about one millimetre of mercury. The frequencies of 29 lines were determined, and it was found that they could be represented by a simple formula from which the frequencies of the remaining lines could be calculated. For seventeen of the lines, the intensities and the half-widths were measured, and it was shown that in each case the strength of the line agreed, within the experimental error of about 5%, with the calculated values. Sufficient data are thus available for the use of this spectrum for a study of pressure broadening.

The pressures used in this work (about 1 mm. Hg) were chosen so that the lines were sufficiently narrow for adequate resolution, but still broad enough to permit accurate measurement of their widths. The line-breadth constants* varied between 2.10^{-4} and 5.10^{-4} cm.$^{-1}$ at a pressure of 0 5 mm. Hg. This pressure lies near the middle of the range of pressures indicated above as suitable for the study of pressure broadening, and it was therefore possible to continue investigation of the spectrum at both higher and lower pressures. In the latter case a complication arises due to the disturbance of thermal equilibrium in the gas; the absorption of energy from the radiation tends to equalize the populations of the upper and lower levels of the transition, and at low pressures the frequency of collisions becomes too small to counteract this process fully and thus maintain the distribution appropriate to the temperature of the gas. The effect has been analysed, and the results, which are in close agreement with the theoretical predictions, will be published shortly in another paper.

The experimental determination of the absorption at higher pressures (up to 60 cm. Hg) had been carried out in 1945 during a preliminary survey of the ammonia spectrum. This survey was mentioned in a précis of the work (Bleaney and Penrose, 1946 a) but was excluded from the more detailed paper (Bleaney and Penrose, 1946 b). The reasons for this were (*a*) the experimental technique was different, (*b*) only a qualitative interpretation of the results was possible before the analysis of the spectrum had been made. In this paper the experimental technique and results are described, and the absorptions at pressures of 10 cm. and 60 cm. Hg are compared with those computed from the measurements on the individual lines at 0·5 mm. pressure. At the high pressures no lines are resolved, and the computation, which involves only simple, though lengthy, numerical work, is based on the assumption that the widths of the lines vary directly as the pressure.

§ 2. THE EXPERIMENTS

2.1. *The method*

The absorption in ammonia at pressures approaching an atmosphere is so large that it may conveniently be measured using path lengths of the order of a metre. The experiments of Cleeton and Williams (1934) showed that radiation of a wave-length of 1·25 cm. decays to half intensity in a distance of about 80 cm. These experiments were carried out by methods similar to those used in the infra-red, using a large mirror to collimate the radiation; the ammonia was contained in a cloth bag which could be introduced between oscillator and detector. For accurate measurements, collimation of the radiation by means of hollow wave-guides is much superior, and a cell of rectangular wave-guide one metre long was used in the experiments now to be described. The ends of the cell were sealed with thin mica windows so that it could be evacuated or filled with ammonia at any desired pressure up to about one atmosphere. Waves of the H_{01} type were excited in the wave-guide by a reflection klystron oscillator, and their intensity was measured by means of a vacuum bolometer placed on the far side of the ammonia cell. The absorption due to the ammonia was determined from the reduction in the intensity registered by the bolometer on admitting gas to the cell. The attenuation due to the gas is greater in a wave-guide than in the unbounded medium by a factor

---

* For a narrow line the line-breadth constant equals half the width of the line at half-intensity; it is thus half of what is generally termed the " line-width " in spectroscopy. (See equation (2) below.)

$$\frac{\mathbf{c}}{v_g}=\left(1-\frac{\lambda^2}{\lambda_c^2}\right)^{-\frac{1}{2}},$$

where v_g is the group velocity of the waves in the guide. The critical wave-length λ_c may be calculated from the dimensions of the wave-guide, and hence the attenuation in the guide can be converted to absorption coefficent in the unbounded medium.

2.2. *The apparatus*

Wave-guide windows. Reflections from the windows of the wave-guide cell cause a drop in the power registered by the bolometer; this would not matter if it remained constant during the experiment. Since the attenuation within the cell is small, interference takes place, however, between the reflections from the windows at either end, and the amount of power reflected depends on the phase difference between the two reflections. This phase difference changes when ammonia is admitted to the cell, because of the small increase in the optical path-length, and the change in the net power reflected by the windows causes an error in the measurement of the absorption. This error is greatest when the change in the optical path-length is about a quarter of a wave-length, as is the case at the highest pressures used in these experiments. Owing to the absorption in the ammonia, the total power reflected by the windows cannot be determined accurately by measurements of the standing wave pattern in the guide preceding the cell; it is therefore important to minimize the error by making the reflection coefficient of each window as small as possible. By use of thin mica (0·002 in. thickness was the smallest that would withstand the pressure of the atmosphere) and careful design of the mounting of the window in the guide, the voltage reflection coefficient was generally kept below 0·1. The corresponding error in the measurement of the attenuation should not exceed ±0·1 decibels.

Oscillator. The source of power was a reflection klystron oscillator of a type designed and constructed in this laboratory by Dr. D. Roaf. To cover the range of these experiments (0·63 to 0·92 cm.$^{-1}$), three tubes were used, and the power available was generally over 10 milliwatts.

Attenuating sections. Matched attenuating sections were inserted (*a*) between the oscillator and test cell, and (*b*) between the test cell and the detector. These ensured that the oscillator worked into a load of constant impedance, and that the source and termination were correctly matched into the wave-guide. A variable attenuator which preceded the bolometer served to adjust the power level.

Detector. The power available at the detector (about 100 microwatts) was sufficient to be measured by a vacuum bolometer, whose sensitive element was a piece of fine steel wire. The bolometer formed one arm of a Wheatstone-bridge circuit. Radio-frequency power was estimated from the out-of-balance current due to the increased bolometer resistance. Preliminary experiments showed that the change in resistance was very nearly, but not exactly, proportional to the incident power. A calibration was therefore made by supplying the bolometer with known amounts of D.C. power. Thus the power measurements were more reliable than they would have been if a crystal rectifier had been employed as detector.

Wave-length measurement. Cylindrical cavities employing the H mode of resonance were used for measuring the wave-length. Three wavemeters of this type covered the range 0·63 to 0·92 cm.$^{-1}$.

2.3. *Experimental procedure*

The absorption was measured at regular intervals of about 0·02 cm.$^{-1}$. The oscillator was set to give the desired frequency, and the power registered by the detector while the cell was empty was observed. This was facilitated by an r.f. switch, consisting of a metal strip which could be inserted into the wave-guide to cut off the detector from the radiation; the galvanometer reading corresponding to zero power in the bolometer could thus be observed at frequent intervals without switching off the oscillator. The latter would be a very undesirable procedure because of the time taken to reach a steady output after any adjustment of the potentials applied to the electrodes of the klystron, and because the tube does not return exactly to the same power output.

Ammonia was then admitted to the wave-guide cell, its pressure being measured on a mercury manometer, and the bolometer reading was taken. This procedure was repeated at a number of different pressures, and the bolometer reading with the cell evacuated, and with zero power, again observed. The zero drift was usually about two to three millimetres, and the change in deflection (about thirty centimetres) for the empty cell was of the same order. A small correction was therefore applied assuming the drifts to have taken place uniformly.

§ 3. RESULTS

At each wave-length, a curve of attenuation in the wave-guide against pressure was drawn; two typical curves are shown in figures 1 and 2, where the reduction to absorption in the unbounded medium has been made to facilitate comparison with the results of other workers, and the pressures (cm. Hg) are plotted as abscissae on two scales differing by a factor 10. Figure 1 shows the results of measurements at a wave-length of 1·25 cm. ($\bar{\nu}=0{\cdot}80$ cm.$^{-1}$), which is close to the centre of gravity of the absorption. The curve rises steeply at low pressures, where the contributions to the absorption from the tails of nearby lines increase rapidly as their widths are increased. Then at high pressures the curve becomes much flatter; this is to be expected, since the intensity at the centre of a pressure-broadened line is independent of the pressure. In this case, the absorption is due to a number of lines, but at sufficiently high pressures, where the widths of the lines are greater than their separation, the attenuation should tend to a constant value.

Figure 2 shows the results at $\lambda = 1{\cdot}47$ cm. ($\bar{\nu} = 0{\cdot}68$ cm.$^{-1}$) which lies on the low-frequency side of most of the strong lines. At pressures of a few cm. Hg, the curve is markedly different from the previous curve, being concave upwards; this corresponds to the fact that the attenuation at a point on the tail of a line should rise with the square of the pressure, as both the number of the molecules per c.c. and the width of the line are increasing. As a number of lines are involved, the curve does not rise as steeply as this, the nearby lines giving a constant contribution when their width becomes greater than the difference between their resonant frequencies and the frequency of measurement, while the contributions from the more distant lines are still rising sharply. At the highest pressures the curve becomes flatter, and should ultimately behave like that for $\bar{\nu} = 0{\cdot}80$ cm.$^{-1}$.

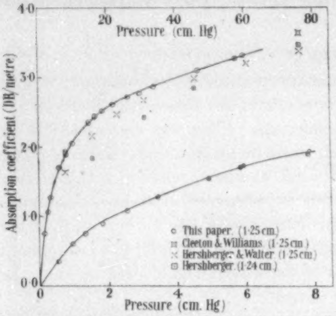


Figure 1. Absorption in ammonia at 1·25 cm. wave-length.

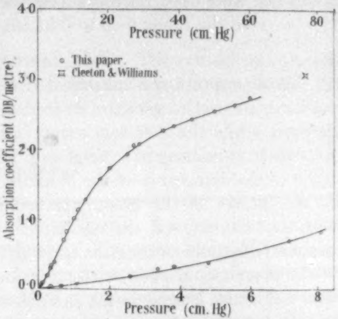


Figure 2. Absorption in ammonia at 1·47 cm. wave-length.

The limiting attenuation will, however, be different since the absorption coefficient is proportional to the square of the frequency.

The results of other experimenters are shown in figures 1 and 2 for comparison. The pioneer work of Cleeton and Williams at atmospheric pressure is represented by a point on each figure; these points appear to be in close agreement with the values predicted from extrapolation of the measurements described in this paper. Recent measurements by Hershberger at wave-lengths of 1·25 cm. and 1·24 cm. are also shown in figure 1, the points being taken from figure 3 of his first paper (Hershberger, 1946) and figure 2 of Walter and Hershberger (1946) respectively. At pressures above 10 cm. Hg, the difference between the absorption coefficients at these two wave-lengths should be less than 3%, as can be seen from figures 4 and 5 of this paper, and as would be expected from the fact that the line-breadth constants at these pressures are 0·1 cm.$^{-1}$ or greater. The large difference between the absorption at these two wave-lengths found by these workers is thus very difficult to understand. As no experimental points are given, nor any estimate of the accuracy, the discrepancy may be attributed to experimental error, and it will be seen that it is of the same order as the difference between Hershberger's values at 1·25 cm. and our curve.

From the graphs of figures 1 and 2, and similar graphs for other wave-lengths, smoothed values for the absorption at a number of selected pressures may be obtained. These absorptions are then reduced to the values appropriate to propagation in an unbounded medium, by multiplying by the ratio of the velocity of light to the group velocity in the guide. The susceptibility of the ammonia gas is so small that it may be neglected in this calculation. The shape of the absorption curves at various pressures may be found by plotting the absorption as a function of the wave-number (figure 3). The pressures chosen were 1, 2, 5, 10, 30 and 60 cm. Hg. The number of points on the curves for the two lowest pressures is inadequate to delineate them accurately, but was sufficient in the preliminary survey of the spectrum to show that a fine structure existed, the separation of whose components must be considerably greater than that suggested by the calculations of Sheng, Barker and Dennison (1941). This fine structure has since been investigated in detail at lower pressures, and these low-pressure curves are not now of particular importance.

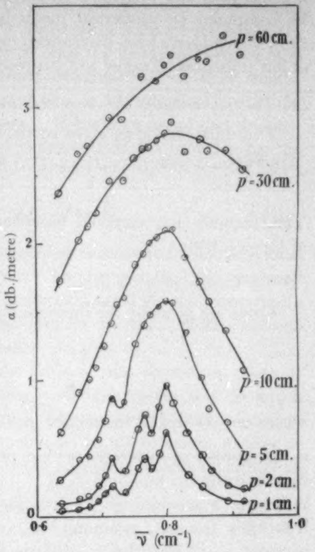


Figure 3. Absorption curves for ammonia at various pressures.

At a pressure of 10 cm.Hg, the lines are sufficiently broad to obscure the fine structure completely, and the absorption curve is quite smooth. The smallness of the scatter of the experimental points is very gratifying, and is well within the possible experimental error. At the high pressures the absorption curves become flatter, and the scatter in the measurements is markedly greater, about ±4% in the curve for 60 cm. pressure. This may be attributed to the small reflections from the windows of the cell, which set up an interference pattern. The introduction of ammonia at 60 cm. pressure changes the optical distance between the windows by more than a quarter wave-length, and the scatter in the points is consistent with the fluctuation of about ±0·1 db. in the power reaching the detector, calculated in the preceding section. The magnitude and sign of the fluctuation change rapidly with wave-length as the cell contains some 60 wave-lengths; thus the scatter in the points on the curve for 60 cm. pressure may be ascribed to this cause, a conclusion that is supported by the smaller scatter at 10 cm. pressure, where the change in the optical path-length is less.

§ 4. DISCUSSION

As mentioned in the Introduction, the main interest of these measurements at high pressures arises in their use for the study of the collision broadening of spectral lines. The complicated nature of the spectrum of ammonia makes the comparison of theory and experiment considerably more intricate than for a single line, and the shape of the absorption curve at high pressures can only be computed by numerical methods. The data upon which the calculation must depend have been determined in the analysis by the authors of the spectrum at low pressures (Bleaney and Penrose, 1946 b). The data required are:

(1) *The frequencies of the component lines of the fine structure*

These are accurately given by the formula

$$=0{\cdot}7935-0{\cdot}0050_5(J^2+J)+0{\cdot}0070_4K^2+0{\cdot}63-\{0{\cdot}0050(J^2+J)+0{\cdot}0070K^2\}^2\ \text{cm.}^{-1}\ (\textit{vacuo}).$$

This formula has received independent confirmation from the experiments of Good (1946).

(2) *The intensities of the lines*

These are given by the theoretical formula

$$\int\frac{\alpha}{\nu^2}d\nu=\frac{8\pi^3N_{JK}}{3ckT}\,|\mu_{JK}|^2, \qquad \ldots\ldots(1)$$

where α = absorption coefficient per cm. of path, $|\mu_{JK}|^2=\mu^2\frac{K^2}{J^2+J}$, and N_{JK} is the number of molecules per c.c. occupying the rotational level characterized by the quantum numbers J, K. The authors have shown that the total intensity at a pressure of 4·5 mm. Hg agrees very closely with that calculated from this formula, assuming the value of $1{\cdot}44\times10^{-18}$ e.s.u. for the dipole moment μ. This value is confirmed by the most recent measurements of the dielectric constant of ammonia (Van Itterbeek and De Clippeleier, 1946), which yield the value of 1·437 e.s.u.

(3) *The shape of a line at high pressure*

Van Vleck and Weisskopf (1945), and Fröhlich (1946), have shown that the collision-broadening theory of Lorentz is inadequate when the width of the line become comparable with the resonant frequency. Their modification of the theory leads to a structure factor *:

$$F(\nu_0,\nu)=\frac{1}{\pi}\left\{\frac{\Delta\nu}{\Delta\nu^2+(\nu_0+\nu)^2}+\frac{\Delta\nu}{\Delta\nu^2+(\nu_0-\nu)^2}\right\}. \qquad \ldots\ldots(2)$$

The absorption coefficient at a frequency ν due to a single line of resonant frequency ν_0 is then given by the expression

$$\alpha=\frac{8\pi^3\nu^2N_{JK}}{3ckT}\,|\mu_{JK}|^2F(\nu_0,\nu). \qquad \ldots\ldots(3)$$

---

* Van Vleck and Weisskopf define a "shape factor" which differs from expression (2) by a factor (ν_0/ν). It seems more logical, however, to exclude this factor from an expression for the shape of the line, and to define the "structure factor" so that $\int_0^\infty F d\nu=1$. The absorption at a particular frequency is then given by multiplying the intensity of the line by the structure factor, the intensity being determined solely by the number of molecules in the two levels between which a transition is taking place and by the probability of such a transition.

(4) *The widths of the lines*

The structure factor $F(\nu_0, \nu)$ involves the line-breadth constant $\Delta\nu$, which for a narrow line equals half the breadth of the line at half intensity. The widths of seventeen lines have been determined by the authors at a pressure of 0·5 mm. Hg (Bleaney and Penrose, 1946 b), the remaining lines not being sufficiently well resolved at this pressure to make an accurate determination of their widths possible. Since the widths vary from line to line, some means of estimating the widths of the unresolved lines is necessary. It has been found that the line-breadth constants of the seventeen lines at a pressure of 0·5 mm. Hg can be expressed by the formula

$$\Delta\bar{\nu}\,(\text{cm.}^{-1}) = 5{\cdot}0 \times 10^{-4}\,\sqrt[3]{K^2/(J^2+J)} \qquad \ldots\ldots(4)$$

For present purposes this formula may be regarded as empirical; its theoretical significance will be discussed in another paper. The deviations of the measured values of the line-breadth constants from those calculated by this formula lie within the experimental error, and (4) may be used for the line-breadth constants of all the lines.

These data are sufficient to calculate the shape of the absorption curve at any pressure, provided some basic assumption is made as to the way in which the line-breadth constant varies with pressure. The simple and obvious assumption is that the line-breadth is directly proportional to the pressure, so long as the latter is not so high that the chance of multiple encounters becomes appreciable. The extension of equation (4) to other pressures is then

$$\Delta\nu\,(\text{cm.}^{-1}) = 1{\cdot}00 \times 10^{-2}\,p_{\text{cm.}}\,\sqrt[3]{K^2/(J^2+J)}. \qquad \ldots\ldots(4\,a)$$

The two pressures chosen for computation were 60 cm. Hg and 10 cm. Hg. The former is the highest pressure at which experiments were made, while the latter is a suitable intermediate pressure where the scatter in the experimental points is small. The absorption at these two pressures varies only slowly with frequency, and the form of the calculated curves can be determined by computation at a reasonably small number of frequencies. The form of equation (3) shows that it is preferable to plot α/ν^2 as a function of the frequency or $\alpha/\bar{\nu}^2$ against the wave-number $\bar{\nu}$, rather than the actual absorption coefficient α, as the presence of the ν^2 term in the latter causes the maximum absorption to shift to higher frequencies when the line becomes broad. This shift is not due to displacement of the resonant frequency and tends to obscure any such effect.

The experimental values of $\alpha/\bar{\nu}^2$ for a pressure of 10 cm. Hg are shown plotted against $\bar{\nu}$ in figure 4. The curve computed from the measurements at 0·5 mm. pressure is drawn as a full line, constructed from points calculated at intervals of 0·05 cm.$^{-1}$ between 0·6 and 1·0 cm.$^{-1}$. It will be seen that the experimental points lie very well on the calculated curve, the deviations being within the estimated experimental error. The following conclusions may therefore be drawn:—

(*a*) There is no doubt that the widths of the lines vary accurately as the first power of the pressure. If the widths varied with the square root of the pressure, as is suggested by some experiments (see Elsasser, 1942), the calculated absorptions would be too small by a factor of $\sqrt{200}$, since the widths have been determined at a pressure of 0·5 mm. Hg and are here extrapolated to a pressure of 10 cm. Hg.

(*b*) There is no significant shift in the resonant frequency of the lines at this pressure.

In figure 5 the experimental values of $\alpha/\bar{\nu}^2$ at a pressure of 60 cm. Hg are shown, together with the calculated curve. There is now an appreciable discrepancy between the measured and the calculated values, the experimental points indicating that the absorption is materially greater than expected at the low-frequency end. The maximum has moved well away from 0·78 cm^{-1} to lower frequencies; no such shift would be predicted by the structure factor. The experimental values do not extend to sufficiently low frequencies to determine the position of the maximum, but in any case the discrepancy between them and the calculated curve is considerably greater than the experimental error. In any attempt to explain the discrepancy, the following possibilities must be considered:

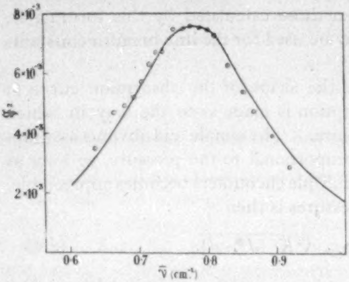


Figure 4. Curve of (absorption coefficient per cm. against wave-number²) at a pressure of 10 cm. Hg.
○ Experimental points.
—— Calculated curve.

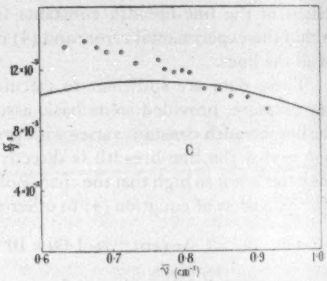


Figure 5. Curve of (absorption coefficient per cm. against wave-number²) at a pressure of 60 cm. Hg.
○ Experimental points.
—— Calculated curve.

(*a*) At high pressures the resonant frequencies of the lines are displaced towards lower frequencies, due to interaction with neighbouring molecules. Since ammonia is a strong dipole, the most important interaction should be a Stark splitting due to the electric field. Polder (1942) has pointed out that only a quadratic effect would be expected in weak fields, the two energy levels of the inversion doublet being given by the formula of Penney (1931),

$$W_{1,2} = \pm\sqrt{\left(\frac{\Delta}{2}\right)^2 + (\mu' E)^2},$$

where

$$\mu' = \mu\frac{MK}{J(J+1)}.$$

Expansion in ascending powers of $(\mu' E)^2$, assuming the latter to be small compared with $(\Delta/2)^2$, gives for the energy difference between the two levels*

$$W_1 - W_2 = \Delta + \tfrac{1}{2}\frac{(\mu' E)^2}{\Delta}. \qquad \ldots\ldots(5)$$

This formula indicates that a displacement towards higher frequencies should occur; if the mean value of E is taken as μ/d^3, where d is the mean distance between the molecules of the gas, the displacement would amount to a few hundredths of a cm.$^{-1}$ at a pressure of 60 cm. Hg. The Stark-effect shift is thus not only of the wrong sign, but also of the wrong magnitude to explain the observed shift.

(*b*) At high pressures the widths of the lines are not directly proportional to the pressure.

At 60 cm. pressure the mean distance between the molecules is about 37 A., while for the broadest lines a collision occurs when the molecules are some 14 A. apart; the chance of multiple encounters is thus appreciable. It is therefore difficult to estimate the sign or magnitude of the deviations from the linear law, but the deviations should be greatest for the widest lines, since these are due to molecules with the largest collision cross-section. These lines lie in the main towards the high-frequency side (the strongest lines, for which $J = K$, all lie at wave-numbers $> 0{\cdot}79$ cm.$^{-1}$); and they are all so broad ($\Delta\bar{\nu} \approx 0{\cdot}5$ to $0{\cdot}6$ cm.$^{-1}$) that the absorption due to them is practically constant over the range of wave-numbers of these measurements. The main effect of a change of the width is therefore to shift the curve bodily up or down without material change of shape. On the other hand, to explain the greater absorption observed on the low-frequency side it would be necessary to postulate that the widths of the narrower lines deviate most from the linear law, the widths increasing less rapidly than the first power of the pressure. It thus appears unlikely that the shape of the observed absorption curve could be explained by a simple deviation from the ($\Delta\bar{\nu} \propto p$) law, though the evidence is not sufficient to rule it out.

(*c*) An appreciable amount of radiation is absorbed by a molecule "during a collision".

When the mean distance between the molecules is comparable with the collision diameter for this absorption, a molecule will spend an appreciable fraction of the total time within regions where the electric field is of the order of 10^6 volts/cm. or more. During this time the energy levels of the molecule must be greatly distorted; there is no simple method of estimating the effect on the absorption, but at high pressures, where multiple encounters are the rule, one might expect to obtain a Debye curve similar to those observed in the case of liquids.

---

* *Note added in proof.* The second-order Stark effect in ammonia has been observed by Coles and Good (1946), who find that the splitting can be represented by the formula

$$\delta\bar{\nu}\ (\text{cm.}^{-1}) = 1{\cdot}5 \times 10^{-4}\,[MK/(J^2+J)]^2 E^2.$$

This is in good agreement with (5), which, on insertion of the numerical values of Δ and μ, gives

$$\delta\bar{\nu}\ (\text{cm.}^{-1}) = 1{\cdot}3 \times 10^{-4}\,[MK/(J^2+J)]^2 E^2.$$

§ 5. CONCLUSION

The inferences which can be drawn from these measurements may be summed up as follows:—

(*a*) The widths of the lines vary linearly with the pressure between 0·5 mm. and 10 cm. Hg.

(*b*) There is no appreciable shift in the resonant frequencies of the lines at pressures up to 10 cm. Hg.

(*c*) The shape of the absorption curve at 10 cm. pressure conforms to that calculated from the structure factor of Van Vleck and Weisskopf, but the measurements do not extend far enough in frequency to provide an adequate test of this factor.

(*d*) At a pressure of 60 cm. Hg, the interactions between the molecules are so large that the absorption curve is appreciably distorted.

It is obvious that considerable extension of the measurements described in this paper is desirable, and has already begun. The frequency range is to be increased so as to include substantially the whole of the extent of the absorption, and to afford an experimental test of the structure factor. The distortion of the absorption curve at 60 cm. Hg should then be more obvious, and will be examined at still higher pressures. Preliminary measurements indicate that the shift in the maximum of the curve of $\alpha/\bar{\nu}^2$ against $\bar{\nu}$ will become even more pronounced at higher pressures. This is indicated by the fact that the curves of absorption against pressure for $\bar{\nu}=0{\cdot}7$ cm.$^{-1}$ begin to turn upwards again at about 90 cm. Hg instead of flattening off, while similar curves for $\bar{\nu}=0{\cdot}9$ cm.$^{-1}$ do not. This could not be the case if the conclusions (*a*) and (*b*) above were still valid at these pressures.

ACKNOWLEDGMENT

The work described in this paper was carried out on behalf of the Director of Physical Research, Admiralty, and the authors wish to record their thanks for permission to publish.

REFERENCES

BLEANEY and PENROSE, 1946 a. *Nature, Lond.*, **157**, 339.

BLEANEY and PENROSE, 1946 b. *Proc. Roy. Soc.*, A, **189**, 358.

CLEETON and WILLIAMS, 1934. *Phys. Rev.*, **45**, 234.

COLES and GOOD, 1946. *Phys. Rev.*, **70**, 979.

ELSASSER, 1942. *Harvard Meterological Studies*, No. 6, pp. 45–48.

FRÖHLICH, 1946. *Nature, Lond.*, **157**, 478.

GOOD, 1946. *Phys. Rev.*, **70**, 213.

HERSHBERGER, 1946. *J. Appl. Phys.*, **17**, 495.

PENNEY, 1931. *Phil. Mag.*, **11**, 602.

POLDER, 1942. *Physica*, **9**, 908.

SHENG, BARKER and DENNISON, 1941. *Phys. Rev.*, **60**, 786.

VAN ITTERBEEK and DE CLIPPELEIER, 1946. *Physica*, **12**, 97.

VAN VLECK and WEISSKOPF, 1945. *Rev. Mod. Phys.*, **17**, 227.

WALTER and HERSHBERGER, 1946. *J. Appl. Phys.*, **17**, 814.

THE FUNDAMENTAL CONCEPTS CONCERNING SURFACE TENSION AND CAPILLARITY

BY R. C. BROWN,
University College, London

MS. received 8 *November* 1946; *lecture delivered* 6 *December* 1946

ABSTRACT. It cannot be denied that our elementary discussions of the principal phenomena connected with surface tension and capillarity are often vague and unsatisfying. The simple statics of the systems dealt with are frequently obscured and circumvented by the introduction of the concept of free surface energy to replace surface tension. The lecture is an attempt to clarify some of the points which arise from this failure to come to grips with fundamentals.

On the basis of the usual idea of cohesion between molecules it can be shown, for instance, that, in contradiction to what is often contended, it is not necessary to deny the reality of surface tension during the course of an explanation of the common phenomena which were, at one time, regarded as providing evidence of its existence. It is possible, also, to gain a less abstract conception of the distinction between free and total surface energy than that provided by a purely thermodynamical discussion.

The customary assumption that in a system containing a solid/liquid interface the surface energy of the solid plays an identical rôle with that of the liquid is criticized, and the conception of surface energy as the work done during the rupture of a column of material is examined. Capillary elevation is regarded as a consequence of negative surface tension in the liquid at the solid/liquid interface, and the usual expression for the capillary rise is derived from this idea..

---

§ 1. THE EXISTENCE OF SURFACE TENSION

IT has become customary in recent years for authors of text-books to deny the existence of surface tension, especially (and somewhat incongruously) when they are presenting an elementary version of Laplace's theory. It is frequently made to appear that the theory explains *away* surface tension instead of accounting for it in a physical manner.

Naturally enough, however, the non-existence of surface tension is not emphasized on those pages which deal with its experimental determination.

Examples of the belief in the non-existence of surface tension are numerous. Thus Newman and Searle (*The General Properties of Matter*) write: "It should be realized, however, that the term 'surface tension' is misleading by reason of its suggestion that there is a real stretching force tangential to the surface of a liquid" (page 163). Champion and Davy (*Properties of Matter*) refer to surface tension as a "useful fiction" (page 99) while Adam (*The Physics and Chemistry of Surfaces*) regards it as a "purely mathematical device" (page 4), and maintains that ".... surface tension does not exist as a physical reality, and is only the mathematical equivalent of the free surface energy" (page 5).

It is clear that we must ask how it comes about that when a soap film supports a weight it should be believed to do so by virtue of a "useful fiction" or a "mathematical device", while a sonometer wire or a clothes line is credited with no such mystical power but is supposed to be in a perfectly real state of tension.

As already mentioned, denials of the reality of surface tension appear to arise during the elementary consideration of the hypothesis which for brevity we shall call "Laplace's theory". It is explained that if we postulate the existence of cohesive forces between molecules of a liquid, then each molecule which is situated in or near the surface is acted upon by a force directed inwards and normal to the surface. It is therefore necessary to do work against this force in order to take a molecule from the interior to the surface, and consequently the surface molecules possess greater energy than those inside the liquid. Therefore, in accordance with the principle that every system moves towards a state of minimum potential energy, if free to do so, the surface of a liquid shows a tendency to contract. It is at this stage that we usually come across a statement to the effect that surface *energy* is therefore the reality and that surface tension is a fiction which is "mathematically equivalent" to surface energy.

There are two points to be mentioned in connection with this argument. In the first place, as with all natural phenomena, our belief in the existence of a tension in the tangent plane of a liquid surface must, of course, be based on experimental evidence, and this is surely convincing. We know that in order to maintain a soap film in equilibrium it must be acted upon by an external force parallel to its surfaces. Figure 1 represents a section of film stretched between wires of circular section A and B. If B is fixed and A is free to move, then in order to maintain equilibrium A must be acted upon by an external force *F*. Therefore, if we take any section of the film at, say C, we conclude that for the equilibrium of AC, the film to the right of C must act on the portion to the left with a force equal and opposite to *F*. This force is found to be proportional to the length of the film perpendicular to the plane of the paper and not to the area of cross-section, which suggests that it has its seat in the surface. In fact, it could not act in the interior of the liquid because this would imply that the pressure in the interior of the film was lower than that in the surrounding atmosphere, and also that a change of shape of the film due to an extension would be opposed by a force acting in the interior, which is contrary to experience, which shows that fluid matter cannot sustain a shear.

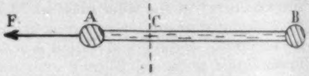


Figure 1.

Secondly, it is erroneous to conclude that surface tension is non-existent merely because a certain elementary development of the postulate of molecular cohesive forces of limited range leads in the first place to an explanation of surface energy. On the contrary, if surface energy exists, it must be necessary to perform work in order to extend a surface, and therefore, when a plane soap film is stretched, there must exist a force parallel to its surface which opposes the stretching and necessitates the application of an external force and the performance of work. It is true that this argument may be regarded as only an indirect way of showing that cohesive forces give rise to surface tension in liquids, but that is a very different matter from denying its existence.

When a liquid surface is extended, the external work is not performed *directly* on the molecules which are brought into the surface from the interior. The system is like a simple machine, the load being the force necessary to transfer the molecules to the surface and the effort being the external force.

In the soap-film example which has just been discussed, the external force balances the surface tension directly, while in the case of the expansion of a bubble or drop by the injection of air or liquid the external force does work against the excess pressure in the interior. We know by a simple argument in statics that this excess pressure implies a tension in the surface. In whatever way the surface of a liquid is extended, we find that the necessity for performing work, which is forecast by Laplace's theory, can be traced to the existence of an opposing force acting tangentially to the surface. The fact is, therefore, that any theory which accounts for surface energy in liquids must, of necessity, also account for surface tension. Free energy per unit area and tension per unit length in a liquid surface are physically equivalent (not merely "mathematically equivalent" as is so often stated). Bearing this in mind, we conclude that it must be possible to develop the original postulates of the theory in such a way as to lead to a more direct explanation of surface tension than is to be found in the usual method of treatment. A simple argument which achieves this is presented in the next paragraph.

Instead of leaving the discussion at the stage where it is shown that the inward cohesive forces acting on the surface layer cause the surface to assume the smallest area compatible with the action of other factors, we should note that when this minimum area has been reached the molecules in the surface are still attracted inwards and possess potential energy with respect to those in the interior. Objects do not remain in positions of high potential energy unless the force by virtue of which they possess that energy is balanced by some other equal and opposite force. Suppose that the surface has reached its minimum area and that the concentration of molecules near the surface is the same as it is in the interior. This condition cannot persist because there is no force to balance the inward attraction experienced by the surface molecules. These will therefore continue to pass into the interior until the reduced concentration in the surface sets up a pressure gradient opposing the inward attraction. When equilibrium is reached, therefore, there are fewer molecules per unit volume in the surface than in the interior and we know from our experience of the effect of pressure on the volume of a liquid that a smaller density means a smaller pressure. Thus the surface layer is in a state of tension compared with the condition of the interior.

It will be noticed that the *fluid* character of the substance is essential to this argument. There are no shearing forces within the liquid itself to oppose the inward flow of molecules until the lack of molecules near the surface establishes an opposing force. The lower molecular concentration in the surface is an example of the general principle that molecules distribute themselves with smaller concentrations at places of higher potential energy.

When equilibrium is established, the potential energy which a molecule would lose by moving from the surface to the interior is equal to the potential energy of strain which the surface layer would gain by this transfer (cf. a small downward displacement of a body supported by a helical spring). In other words no molecules have any resultant force acting upon them, otherwise they are not in equilibrium.

A deeper analysis of the equilibrium between the surface and the interior has been worked out by Bakker (Vol. VI, Wien-Harms' *Handbuch der Experimentalphysik*) and it seems worth while to give a brief outline of this treatment.

Imagine a small plane surface of area Δs drawn anywhere in a liquid and let F be the force with which the liquid on one side of the plane attracts that on the other side. Then the *cohesion*, K, is given by

$$K = F/\Delta s.$$

Cohesion is also known as *intrinsic pressure*, and is an extremely large pressure of the order of 10000 atmospheres or more.

In addition to K there is, in general, another force which also presses together the two portions of liquid separated by Δs. This is, of course, what we ordinarily know as the pressure in the liquid, a quantity which, unlike K, is measurable directly by the force on a piston, one side of which is exposed to the liquid. It is produced by agents external to the liquid itself such as gravity, the atmosphere, artificial compression in a cylinder, or simply by the vapour pressure if the liquid partially fills a closed vessel.

If the value of the pressure is p at any point where the cohesion is K, then the *total pressure*, P, at this point is given by

$$P = K + p.$$

Thus the force per unit area with which the liquid on one side of an imaginary element of area presses against that on the other side is made up of the force due to cohesion and the force due to external causes. In other words, the condition of compression of the liquid is the same as it would be if no cohesion existed but a pressure equal to K were applied externally in addition to p. It is necessary to use two different terms to distinguish between P and p, and we shall adopt *total pressure* for the former and simply *pressure* for the latter.

The crux of the argument which demonstrates that cohesion gives rise to surface tension is the fact that the *total pressure* at a point in a fluid has the same value in all directions. Although we are accustomed to regarding this as being a property of the ordinary pressure p, yet this is actually a particular case which is true only at points where K has the same value in all directions.

In the elementary proof that the pressure at a point in a fluid is independent of the orientation of the plane across which it is measured we consider the action of the external forces due to p upon the three faces of a prismatic element of fluid. In fact, however, the total force per unit area on any face is not p but $K+p$, i.e. P. It is this force which would have to be applied to the faces of the element (by solid pistons for example) in order to maintain its original condition if it were isolated from the body of the liquid. Thus the property of being independent of direction is fundamental to P rather than to p. We shall see that at any point in the surface layer K is a function of direction so that p is also dependent upon direction in this region and it is this which gives rise to surface tension.

Suppose that the attractive force between two small elements of the liquid becomes negligible when the distance between them exceeds a certain small value c. Then if we consider a quantity of liquid in the form of a column of uniform cross section Δs it is clear that the force of attraction between the two parts into which any cross section divides the column is independent of the position of the section provided that its distance from either end of the column is not less than c. Therefore a surface layer in which K is less than it is in the interior extends inwards from each of the end faces of the column. If the substance is supposed to be inextensible the depth of this layer is equal to c.

Figure 2 represents a portion of the surface layer of a liquid, that is to say, the region in which a particle of liquid experiences a net inward attraction. Let a small area Δs be drawn in the surface layer parallel to the surface.

Then, if at the level of Δs the ordinary fluid pressure taken perpendicular to Δs is p, and if K_n is the cohesion taken in the same direction, we can write

$$P = K_n + p, \qquad \ldots\ldots(1)$$

where P is the total pressure at Δs.

Since K_n is the attractive force per unit area between the two portions into which the cross-section Δs divides the vertical column indicated in figure 2, the value of K_n must be zero at the outer boundary of the surface layer (if we suppose that the vapour exerts no attractive force on the liquid) and it increases as Δs moves into the liquid, until at the inner boundary of the layer its value is equal to the cohesion in the interior of the liquid.

We shall ignore the effect of gravity upon conditions in the surface layer, and regard p as being entirely due to the atmosphere or vapour above the liquid surface. By considering the equilibrium of the column lying between Δs and the outer boundary of the surface in figure 2 we observe that, in the liquid, the pressure p taken across an element of area parallel to the surface is everywhere equal to the pressure in the medium above the surface.

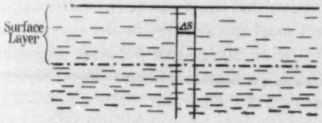


Figure 2.

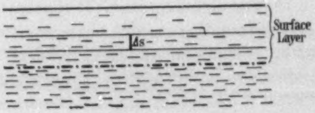


Figure 3.

If we now take Δs at the same level in the surface layer as previously, but perpendicular to the surface (figure 3), we have

$$P = K_s + p_s \qquad \ldots\ldots(2)$$

where K_s and p_s are respectively the cohesion and pressure taken across a small area perpendicular to the surface at this particular depth. Since P is the same in equations (1) and (2), we have

$$p - p_s = K_s - K_n.$$

Now, when Δs is perpendicular to the surface (figure 3), the columns of attracting liquid extend to a distance greater than c on each side of it, whereas when it is parallel to the surface (figure 2) the column on the side nearer the surface is shorter than c, and the attractive force across Δs is therefore less. Thus it follows that for any given point in the surface layer

$$K_s > K_n$$

so that

$$p_s < p,$$

which implies a state of tension parallel to the surface.

The expression for the surface tension may be derived as follows:—In figure 4 some liquid is contained between three vertical walls of a closed vessel A and a movable partition B attached to a rod which passes through the wall of A. It is convenient to suppose that the angle of contact between the liquid and B is 90°, and this may be done by choosing the appropriate solid without affecting the value of the surface tension of the liquid. If the effect of gravity is left out of consideration, the pressure is uniform throughout the interior of the liquid and is equal to that in the gas and/or vapour confined in A. At all points in the surface layer, however, the pressure p_s across an area perpendicular to the surface is less than that in the vapour. Apart from the effect of the surface layer, the partition B would experience no resultant force, but if the layer has a thickness t and unit width perpendicular to the plane of the paper, the force on B towards the left is equal to

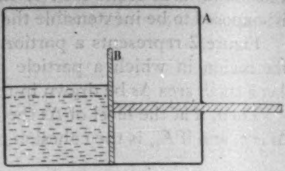


Figure 4.

$$(p-p_s)t,$$

where p_s now represents the *mean* pressure in the surface layer in the direction parallel to the surface. The surface tension (γ), as commonly understood, is equal to the external force which must be applied to the rod in order to keep B in equilibrium and is therefore given by

$$\gamma=(p-p_s)t. \qquad \ldots\ldots(3)$$

At first sight, equation (3) appears to suggest that the pressure (p) in the liquid and surrounding gaseous phase has a direct mechanical effect on the magnitude of γ. This is not the case, however, because an increase of p causes an equal increase to occur in P and therefore also in p_s (equations (1) and (2)). Nevertheless it is worth noting that the *actual* tension in the surface is not γ but $-p_s t$, to which we may give the symbol σ.

As already mentioned, the belief in the non-existence of surface tension appears to arise from the misconception that cohesion in a liquid explains only surface energy, and it is hoped that the foregoing demonstrates that this is, in fact, a misconception. At the same time the author does not admit that surface tension owes its existence to any *theory*, but rather to the simple statics of, say, a plane soap film or bubble.

In the course of a very brief reference to Bakker's treatment, Adam (*The Physics and Chemistry of Surfaces*, p. 4) admits that it is "mathematically correct" but makes the following criticism:—"It is certain now, however, that the surface layer is ordinarily only a few molecules thick, so that the conception of pressure and its variation parallel to the surface becomes rather intangible in terms of molecules." The answer to this is, surely, that we do not regard the surface layer as a collection of stationary molecules but as a *region* within which the molecules experience resultant forces which are absent in the interior. Even if the region is very thin, there is within it an infinite number of possible levels which can be occupied by the molecules instantaneously during their thermal agitation. A different value of P is associated with each level, and over a sufficiently long period of time each level will contain the centres of a definite average number of molecules.

§ 2. A MECHANICAL ANALOGY

A mechanical model which represents the mechanism of surface tension can be devised as follows. In figure 5, ABCD is a continuous length of cord. The portions AB and CD rest on smooth horizontal shelves at different levels, and pulleys are provided at B and C. The portion AB represents a collection of molecules in the interior of a liquid while CD represents molecules in the surface. We can regard CD as a kind of average level between the inner and outer boundaries of the surface layer of the liquid. The path which the molecules take in moving from the interior to the surface is represented by BC. The molecular attraction against which the molecules of the liquid must move in order to reach the surface is replaced in the model by the earth's gravitational field acting on every element of the cord between B and C, the total force being the weight of BC. In order to maintain the system in equilibrium it is necessary to apply at D an external force (γ) equal to the tension in CD which, in turn, is equal to the weight of BC. If the cord is inextensible and D is caused to move to the right, the work done by γ appears as the gravitational potential energy gained by the length of cord which is raised to the higher level. If the cord is extensible, however, a given amount of external work performed by γ raises a smaller weight of cord than otherwise and the remainder of the work appears as energy of strain because the portion of cord which has been raised has also been extended. If the elastic and thermal properties of the cord are such that it cools when stretched adiabatically it will be necessary to supply heat when the cord is drawn up to the higher level isothermally, so that its behaviour is comparable to that of a liquid surface in this respect also.

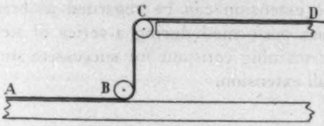


Figure 5.

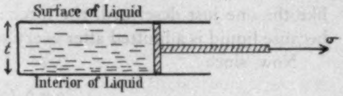


Figure 6.

§ 3. TOTAL AND FREE SURFACE ENERGY

The well known relation between total surface energy (E) and free or available surface energy (γ), viz.

$$E = \gamma - T\frac{\partial \gamma}{\partial T},$$

is derived from the laws of thermodynamics which are independent of the nature of the working substance and of the processes involved. The equation would apply to any system exerting an external force γ which is a function of temperature. Consequently it frequently happens that little more is said about this aspect of the properties of liquid surfaces than that $\partial\gamma/\partial T$ is negative for the great majority of substances, so that when a surface is extended isothermally by unit area it is necessary to impart to the liquid an amount of heat (viz. $-T\partial\gamma/\partial T$). It is quite possible, however, to gain an idea of the mechanism of surface tension which the equation reflects by considering a "simplified" liquid. Imagine that the surface layer of a liquid is contained in a vessel as shown in figure 6 and that its width perpendicular to the plane of the paper is unity. We can suppose that the lower wall of the vessel separates the surface layer from the interior but does not modify the cohesive forces acting across it. In order to maintain the piston in equilibrium it is necessary to act upon it with an external force directed outwards and numerically equal to σ $(=-p_s t)$, the total tension in the surface layer. This is the *total* external force necessary to keep the surface in equilibrium. In the case of a plane soap film, σ is the resultant of an applied force γ acting outwards and the force pt due to the atmosphere acting inwards (equation (3)).

When the piston is pulled out reversibly through an infinitesimal distance, the pressure is slightly decreased everywhere, and if a hole were now made in the lower wall of the vessel a small amount of liquid from the interior would flow into the surface layer and restore the pressure to its original value. This is a useful way of picturing the mechanism of surface extension. The total external work performed during a finite extension can be regarded as being the sum of all the small amounts of work performed during a series of steps like the one just described, the force σ remaining constant for successive steps because liquid is admitted after every small extension.

Now, since

$$\sigma=p_s t=(K_s-P)t, \qquad \ldots\ldots(4)$$

it is evident that we can regard the force σ which the surface exerts on the piston as being the resultant of a force due to cohesion, $K_s t$, and an opposite force Pt tending to expand the surface. In order to extend the surface by a small area, da, it is necessary, therefore, to do a total amount of work $K_s t\,da$ against cohesion. Of this total amount a quantity of work σda is performed by the external force which balances σ, leaving, by equation (4), $Pt\,da$ units of work to be performed by P against the cohesion. Thus the total pressure, P, assists the extension of the surface by its tendency to force the molecules apart. In performing this work the molecules will cool in the same way as an ideal gas cools when it does work and, for an isothermal extension of the surface, an amount of heat equivalent to $Pt\,da$ units of work must be given to the system.

Thus for unit area of surface we can write:

Total surface energy $=K_s t$.
Free surface energy $=\sigma$.
Superficial "latent heat" $=Pt$.

Suppose that we regard P as being due to the thermal agitation of the molecules and assume that it is proportional to the absolute temperature when the separation of the molecules remains constant, i.e. when the layer contained in the vessel in figure 6 is heated at constant volume. This assumption is in accordance with the ideas underlying an equation of state such as that of Van der

Waals in which the total pressure $(p+a/v^2)$ is equated to $RT/(v-b)$. On this assumption

$$\left(\frac{\partial P}{\partial T}\right)_v=\frac{P}{T},$$

and the superficial latent heat, Pt, can be written as

$$T\left(\frac{\partial P}{\partial T}\right)_v t.$$

We can, furthermore, reasonably suppose that at constant volume, since the separation of the molecules is constant, the value of K_s is independent of temperature, so that by differentiating equation (4) we obtain

$$\begin{aligned}\left(\frac{\partial P}{\partial T}\right)_v t&=\left(\frac{\partial K_s}{\partial T}\right)_v t-\left(\frac{\partial \sigma}{\partial T}\right)_v\\ &=-\left(\frac{\partial \sigma}{\partial T}\right)_v.\end{aligned}$$

The superficial latent heat therefore becomes

$$-T\left(\frac{\partial \sigma}{\partial T}\right)_v,$$

and we can replace equation (4), i.e.

$$K_s t=\sigma+Pt,$$

by the energy equation

$$E=\sigma-T\left(\frac{\partial \sigma}{\partial T}\right)_v,$$

where the symbol E replaces $K_s t$ to represent the total surface energy per unit area.

Thus, using very simple ideas, it is possible to establish an equation which is similar to the equation of free energy. The two are not identical because as we have seen, σ is not the same quantity as the surface tension γ, their difference being equal to the product of the pressure p by the thickness of the surface layer t. The two quantities become identical only when the pressure is zero. Even at atmospheric pressure, however, the difference between them is very small, being of the order of $10^6\times10^{-7}$ or 0·1 dyne cm.$^{-1}$ if the thickness of the surface layer is taken to be of the order of a few molecular diameters. Under ordinary conditions, therefore, we can feel justified in regarding the relationship between the free and total energies of a liquid surface as being due almost entirely to the mechanism upon which we have based the derivation of the last equation.

One point of interest in this connection is that we have identified the cohesive force $K_s t$ with the total surface energy, and have assumed that K_s remains constant when the temperature of the surface layer is changed at constant volume. In actual fact the total surface energy, given by $\gamma-T\left(\frac{\partial \gamma}{\partial T}\right)$, does vary with temperature, and two observations can be made concerning this discrepancy. In the first place, in the thermodynamical derivation of this expression for the total energy, the volume of the surface layer is *not* supposed to be constant as the temperature is varied—nor, indeed, is it constant in reality. In these circumstances the cohesion K_s will vary with temperature on account of the changing average distance between the molecules. Secondly, the increased concentration of molecules in the vapour following a rise of temperature will, in practice, diminish the surface tension by reducing the inward attractive force acting on the surface molecules. This last effect, which eventually reduces γ to zero at the critical temperature, is not taken into account in the above discussion, where the only influence of temperature upon σ is supposed to be that which is due to the variation of P with temperature at constant volume. Nevertheless, the foregoing ideas are supported by the experimental fact that at ordinary temperatures, when the effect of the vapour is small, the total surface energy actually varies very little with the temperature, and in some cases is remarkably constant. One example is mercury, as might be expected on account of its low vapour pressure and comparatively low expansion coefficient. It will be realized that constant total surface energy implies a constant value of $d\gamma/dT$, i.e. a linear relationship between surface tension and temperature.

§ 4. THE SURFACE ENERGY OF SOLIDS

The elementary argument which shows that cohesion between molecules leads to surface energy is, as usually presented, equally applicable to both liquids and solids. In both cases the necessary condition is present, namely, that surface molecules are attracted towards the interior. Or, to put the matter in another way, it is necessary to perform mechanical work in order to break a column of cohering matter so as to form two fresh surfaces, whether the substance be solid or liquid. It is frequently stated therefore that, since the step from surface energy to surface tension is "purely mathematical", solids as well as liquids must possess surface tension. It is, however, worth while questioning this argument, if for no other purpose than to discover how it comes about that, while the surface tension of a liquid can be directly determined in a variety of ways commonly applicable to the measurement of a force, the surface tension of a solid is somewhat elusive.

The important point in this connection is, of course, that the spontaneous contraction of a liquid film, which can be prevented only by the application of an external force equal and opposite to the surface tension, has no counterpart in the case of a solid sheet. The essence of the process is the freedom with which the molecules of a liquid can pass from the surface to the interior, that is to say, the inability of a liquid to sustain a shearing stress and to oppose permanently the attempt on the part of the surface tension to change its shape. In a solid the cohesive forces acting inwards on the surface molecules are balanced by elastic reactions set up within the solid itself so that it is unnecessary to apply an external force in order to prevent the contraction of the surface.

It cannot be said, therefore, that a solid possesses surface tension in the sense in which the term is applied to liquids. Many authors, nevertheless, use the terms "surface energy" and "surface tension" indiscriminately in connection with solids—frequently having denied the existence of surface tension elsewhere in their writings.

A good deal of light can be thrown on the nature of surface energy and the distinction between the properties of solids and liquids by examining the fundamental process of the creation of new surfaces by the rupture of a column of matter.

In the first place let it be assumed that the column is composed of extensible material. Supposing that the molecules are in a static condition and are all equally spaced, then in order to effect a rupture it is necessary to create at one cross section an axial separation of molecules equal to c, the maximum molecular separation for which finite cohesion exists. Theoretically this can be brought about by the application of equal and opposite external forces to the molecules lying immediately on each side of a given transverse section. If the forces are applied elsewhere (e.g. at the free ends of the original column) some of the work which they perform will create a state of strain in the material and it would be necessary to subtract the energy corresponding to this in order to arrive at the amount of work required to produce the new surfaces.

In figure 7, AB and CD represent two contiguous layers of molecules to which external forces are applied. As the separation of the layers proceeds, the value of the cohesion K_n parallel to the axis of the column, *i.e.* perpendicular to the surfaces about to be formed, diminishes not only between AB and CD but also across a plane such as EF situated near to CD because there are now fewer molecules within the distance c from this plane on the left-hand side. This effect extends only for a limited distance on each side of AB and CD since the range of molecular attraction is limited.

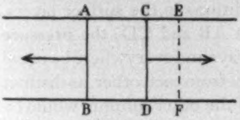


Figure 7.

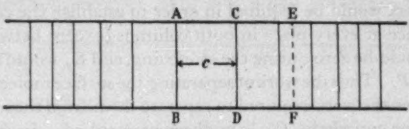


Figure 8.

It is, of course, quite possible that the rupture of a column may not be a completely reversible process. At a certain stage during the stretching, the molecules in the regions which will eventually form the surface layers may become unstable and acquire accelerations which cannot be controlled by adjusting the magnitude of the external forces. The final static state of the substance when rupture is just complete can, however, be discussed without reference to this possibility. In figure 8 the distance between AB and CD has reached the value c. Further separation does not require the expenditure of work, neither does it change the conditions at a plane such as EF. The two parts of the column are now furnished with stable surface layers. If we suppose that the external pressure is zero, then K_n and P (the total pressure) must be equal to each other at every transverse section. Therefore, in the surface layers, e.g. across EF, where K_n is small, P is less than it is in the interior and this implies a greater molecular separation. Thus the condition is as shown diagrammatically in figure 8 where the distances between the transverse lines represent distances between molecules in the direction parallel to the axis of the columns.

The moving apart of the molecules which has occurred as a result of the rupture has, of course, been assisted by the action of P, and the work done by this repulsion must be supplied in the form of heat if the process is to be isothermal. It would therefore appear that when we calculate (as in Laplace's theory) the work done against molecular attraction during the rupture of a column we are in fact finding an expression for the *total* rather than the *free* surface energy, inasmuch as the assistance given by P is ignored.

If we are dealing with a liquid column contained in a tube—for simplicity the angle of contact can be supposed to be 90°—then, as already explained, the diminshed value of K_n in the surface layer sets up a tension parallel to the surface which is capable of doing external work. Superficial latent heat is also accounted for because during a contraction of the surface the closer approach of the molecules to each other as they leave the surface and enter the interior is accompanied by the emission of heat which is the thermal equivalent of the work done during this process by the cohesion between them.

Solid surfaces, however, show no tendency to contract and require no external force to keep them in equilibrium. Nevertheless there is a condition of strain in the surface layer such as that depicted in figure 8, and if this is partially destroyed by, say, allowing a liquid to spread over the solid, thus increasing the value of K_n in the surface layer, some of the heat energy which was required for the establishment of the state of strain will be liberated. The question as to whether mechanical energy also would be derivable *directly* from the disappearance of the strain in the surface depends upon the character of the original process of rupture. If this were completely reversible in every respect no mechanical work would be required in order to establish the conditions in the surface layers, since at every place in both columns (except between AB and CD) the pressure would be zero during the stretching, and K_n would always and everywhere be equal to P. Thus the work of separating the surface molecules from each other, as distinct from the work required to separate them from those of the other column, would be done entirely by P, i.e. by the heat supplied. In this case, therefore, the condition of strain in the surface layer itself would provide no mechanical energy to assist the spreading of a liquid. On the other hand, if the process of rupture is not completely reversible, the solid surface might itself possess available potential energy and therefore exert a tangential force on a spreading liquid. A possible mechanism by which this could be achieved is suggested later.

It is interesting to enquire how the above argument is modified if the ruptured column consists of an inextensible material (supposing that such could exist). In this case all the work is done *after* rupture, because a movement of the external forces before rupture would be contrary to the hypothesis of inextensibility. The work of separation will be completed when the width of the gap between the two new surfaces reaches the value c. The surface layers cannot be in a state of strain if the substance is inextensible, and the mutual repulsion of the molecules (P) cannot assist in the process of rupture because this is complete as soon as the distance between the molecules on either side of the gap exceeds its original value by only an infinitesimal amount. Superficial latent heat depends upon the existence of compressibility and therefore is zero in the case we are considering. If the substance is a liquid, surface tension exists, however, because an argument such as Bakker's is independent of whether or not the liquid is compressible. The work done in separating the liquid columns after rupture appears as surface energy. If the newly formed liquid surfaces are allowed to do work by contraction, the work derivable when the two columns of reduced cross-section are allowed to come together again will fall short of the original work of separation by the amount of work done by surface tension during the contraction of the surfaces. In other words, the attraction between the columns will have been decreased in proportion to the decrease of the area of the surfaces facing each other.

Once again the reasoning as to the existence of surface tension is dependent upon the fluid character of the substance and, therefore, does not apply to a solid. The only way in which the energy of the two separated columns of an inextensible solid can be recovered is by allowing them to come together again, and it seems hardly appropriate to associate such potential energy of separation with the *surfaces*. The newly formed surface layers may acquire stresses as a result of the removal of one column from the other but a stressed inextensible solid is incapable of performing work. The case of an inextensible solid is included here because, although it is hypothetical, it is shown in the next section that it is possible to create a very simple picture of the mechanism of spreading and capillary rise by treating the solid as inextensible.

§ 5. THE EQUILIBRIUM BETWEEN SOLID AND LIQUID SURFACES

In figure 9 the full lines represent the diagram which usually accompanies an investigation of the equilibrium between solid and liquid surfaces. By considering a displacement in the form of a spreading of the liquid over the solid surface and remembering that the condition for equilibrium is that a small displacement requires no external work, we write

$$\gamma_L \cos\theta = \gamma_S - \gamma_{SL}, \qquad \ldots\ldots(5)$$

where γ_L and γ_S are respectively the free surface energies of the liquid and solid against vapour and γ_{SL} is the free energy of the composite surface between solid and liquid. The dotted lines are inserted in figure 9 in order to indicate a less abstract state of affairs than that suggested by the unfinished diagram. We are concerned only with surface energies so that the effects of gravity can be ignored. The angle between the two solid surfaces is arranged so that the liquid/vapour surface is plane, and therefore the pressure in the liquid is equal to that in the vapour. The spreading can therefore be effected without the performance of work by injecting more liquid from outside through a tube which ends in the interior of the existing mass. This elucidation is, however, purely incidental.

The foregoing treatment is physically somewhat unsatisfying. The relation derived depends only upon the definition and conservation of free surface energy and therefore gives little information beyond the fact that there is, in general, an equilibrium value for the contact angle between a given solid and liquid. It is tempting to replace the surface energies by surface tensions and to regard the equation as an expression of a state of equilibrium between the three tensions acting at A (figure 9). Even cursory reflection at this stage, however, gives rise to doubts as to the justification for supposing that the solid/vapour surface assists spreading in the same way as would a liquid surface, i.e. by contracting and exerting a tension on A while doing so. Nevertheless the matter is usually left in this state.

In the following treatment it is assumed that the solid surface has no free energy, or more precisely, that its free energy is not modified when liquid replaces vapour in contact with the surface. We deal, therefore, with the inextensible solid discussed in the previous section and show that the existence of spreading and capillary rise can be explained on this basis. This is equivalent to saying that the occurrence of these phenomena does not in itself constitute any experimental evidence as to the existence of free energy in a solid surface.

Outside a solid surface there is a field of force directed perpendicularly inwards and extending beyond the surface for a distance equal to the maximum range of attraction of the solid molecules. When a liquid is in contact with such a surface this field will diminish or overcome the attraction towards the interior of the liquid which the surface molecules of the liquid experience and which normally gives rise to the tension in the liquid/vapour surface. The surface tension ($\gamma_{L(S)}$) of the liquid surface in contact with the solid is therefore less than that of the liquid/vapour surface and may, quite possibly, be negative, i.e. a surface pressure. Since we have assumed that the state of the solid surface is not modified by the presence of liquid in contact with it, it follows that the attractive force of the solid on a liquid particle at A (figure 9) must be normal to the surface and have no tangential component. Therefore the equilibrium of A is represented by the equation

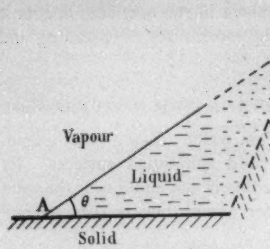


Figure 9.

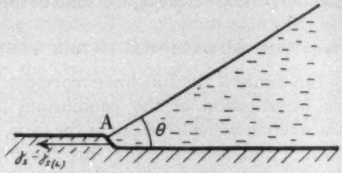


Figure 10.

$$\gamma_L \cos\theta + \gamma_{L(S)} = 0. \qquad \ldots\ldots(6)$$

Equation (6) shows that $\gamma_{L(S)}$ is positive when θ exceeds 90°, zero when θ is equal to 90°, and negative when θ is less than 90°. In the last case the liquid surface in contact with the solid is in a state of pressure tending to make the liquid spread.

If, as is quite possible, the solid and liquid are such that

$$\gamma_L < -\gamma_{L(S)},$$

equilibrium is not possible according to equation (6). The force opposing spreading ($\gamma_L \cos\theta$) is insufficient to balance the spreading force ($-\gamma_{L(S)}$) even when θ has become zero. The liquid therefore spreads until it forms a layer of molecular thickness in which, ultimately, there is no tension or pressure. The case of zero contact angle must not be considered as a special one such as would be represented by any given *finite* value of θ, but rather as being characteristic of the *class* of substances for which $\gamma_L \leqslant -\gamma_{L(S)}$.

The mechanism of spreading implied in this treatment is simply the creation of a surface pressure by the attraction of liquid molecules into the surface layer by the field of force of the solid. This is equivalent to the "squeezing-out" process envisaged by Leslie, the attraction of the solid upon liquid molecules beyond the first layer causing a tangential pressure in the surface layer. This idea is not popular among present-day surface specialists, but it must evidently be considered on the same footing as the simple theory which ascribes the free energy of a liquid/vapour surface to an inward attraction acting on the surface molecules. If an outward attraction is superimposed the result will be a reduced or negative surface energy.

If we no longer suppose that the free energy of the solid surface is unchanged when liquid spreads over it, then as a result of the replacement of vapour by liquid in contact with the solid surface, a certain amount of mechanical energy becomes available and the principle of virtual work leads to the equation

$$\gamma_{\mathrm{L}} \cos \theta = \gamma_{\mathrm{S}} - \gamma_{\mathrm{S(L)}} - \gamma_{\mathrm{L(S)}}, \qquad \ldots\ldots(7)$$

where $\gamma_{\mathrm{S(L)}}$ is the free energy of unit area of the solid surface when covered with liquid. This is, of course, identical with equation (5), the energy of the composite surface (γ_{SL}) being expressed as the sum of the energies of the two surfaces of which it is composed. Since $\gamma_{\mathrm{L}} \cos \theta$ and $\gamma_{\mathrm{L(S)}}$ in equation (7) are real forces it follows that $\gamma_{\mathrm{S}} - \gamma_{\mathrm{S(L)}}$ must represent a real force acting tangentially on A, and it is necessary to seek a mechanism by which this force can be exerted on the liquid.

The solid can no longer be regarded as inextensible if the condition of its surface is modified by the nature of medium with which it is in contact. The extension in a direction normal to the surface, which, as already explained, is produced when the solid surface is formed by breaking a column, will be reduced by the presence of the liquid because the value of K_n in the surface layer of the solid will be increased, and for equilibrium this necessitates an increased value of P. There will, therefore, be a very small but definite "step" or change of level in the solid surface at A, as shown in figure 10, and the attraction which the solid exerts on a liquid particle at A is no longer normal to the general direction of the solid surface as it would be in the absence of the step. The tangential component of this attraction must, by equation (7), be equal to $\gamma_{\mathrm{S}} - \gamma_{\mathrm{S(L)}}$ and be directed away from the liquid.

It has previously been suggested in this lecture that the existence of free surface energy in a solid would indicate that the process of formation of the surface by the rupture of a column is irreversible. In these circumstances the wetting of the surface by a liquid would also be irreversible. This is a possible explanation of the well known observation that the advancing and receding contact angles differ considerably from each other.

§ 6. CAPILLARY RISE

The mechanism of capillary rise is, of course, closely related to that of spreading. It is simply spreading in a vertical direction which is eventually arrested by the force of gravity acting on the raised liquid.

A simple treatment of capillary rise can be presented by assuming, in the first place, that the walls of the tube or plates between which the liquid rises do not, themselves, exert a tangential force on the line of contact of the meniscus with them. That is to say, the surface energy of the solid is unaltered by wetting and can therefore be treated as inextensible. In this case the driving force is the surface pressure in the layer of liquid adjacent to the walls and equilibrium is established when the upward force due to this pressure is equal to the weight of the raised liquid column. Figure 11 is a representation of this state of affairs. The thick lines indicate surfaces which are in a state of tension and the dots show surfaces in a state of pressure. Inside the tube the surface pressure $-\gamma_{L(S)}$ supports the surface layer which covers the meniscus and this in turn supports the weight of the raised column by cohesion. Thus considering the equilibrium of the raised liquid and ignoring the weight of the liquid in the meniscus, we have

$$-2\pi a\gamma_{L(S)} = \pi a^2 hg\rho, \qquad \ldots\ldots(8)$$

where a is the radius of the tube, h is the height of the raised column and ρ is the density of the liquid. Furthermore, assuming (as is always done) that the magnitude of the angle of contact is independent of the action of gravity, we have, either by equation (6) or by considering the equilibrium of a liquid particle at the extreme edge of the meniscus,

$$-\gamma_{L(S)} = \gamma_L \cos\theta. \qquad \ldots\ldots(9)$$

Combining the last two equations gives the usual relation

$$\gamma_L \cos\theta = \frac{ahg\rho}{2}. \qquad \ldots\ldots(10)$$

If a very simple analogy may be permitted, the surface of the meniscus can be likened to a clothes line. The sum of the two vertical components of the tensions at the end of the line is equal to the weight of the clothes. This is, in fact, how the matter is very frequently treated in elementary text-books: the force $2\pi a\gamma_L \cos\theta$ acting upwards round the edge of the meniscus is equated to the weight of the column. This treatment is unsatisfying, however, unless it includes an explanation of how the edge of the surface layer of the meniscus (i.e. the ends of the clothes line) is supported. The foregoing argument shows that the surface pressure $(-\gamma_{L(S)})$ provides the support.

The case of zero contact angle must next be considered. It has already been explained that when $\gamma_L < -\gamma_{L(S)}$ the contact angle is zero and that for equilibrium a thin layer of liquid must cover all the exposed surface of the solid. Equation (9) is, of course, inoperative in these circumstances and so also is equation (8), because if it were applied as it stands, it would give different values of h for a given liquid with tubes of different materials ($\gamma_{L(S)}$ being dependent upon the nature of the solid), which is contrary to experience when θ is zero. It must be remembered, however, that the surface pressure acts upon the liquid column via the surface of the meniscus, so that, when the tension in this surface (γ_L) is less than $-\gamma_{L(S)}$, it is γ_L which decides the value of h (cf. the fact that the maximum weight which a clothes line can support is governed by its breaking tension). Thus the liquid initially rises to its equilibrium level given by

$$\gamma_L = \frac{ahg\rho}{2},$$

and after this the pressure in the layer adjacent to the walls causes the surface of the meniscus to extend indefinitely up the walls of the tube until, theoretically at any rate, the thin film of liquid so formed meets a similar film spreading up the outside wall of the tube from the liquid surrounding it. In the case we are considering, conditions are favourable to spreading, which means that the film of liquid on the walls of the tube above the meniscus is exerting a net force tending to extend its boundaries. Let this be ϕ per unit length of boundary. Then the equilibrium of a small element of liquid situated at the edge of the meniscus is given by

$$-\gamma_{L(S)}-\phi=\gamma_L. \qquad \ldots\ldots(11)$$

As to the value of ϕ, we can say that if the film is thick enough for its liquid/vapour surface to be beyond the influence of the attractive field of the solid, then the tension in this surface will be γ_L and the pressure in the liquid/solid surface of the film will be $-\gamma_{L(S)}$. Hence the net force which such a film exerts on unit length of its boundary will be the resultant of the pressure and tension in its two surfaces and will be given by

$$\phi=-\gamma_{S(S)}-\gamma_L, \qquad \ldots\ldots(12)$$

which is identical with equation (11). It is not contended that the film is necessarily thick enough to possess two independent surfaces—merely that the condition for equilibrium (equation (11)) requires that the value of ϕ shall be the same as it would be if the film were of this nature.

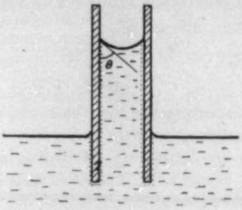


Figure 11.

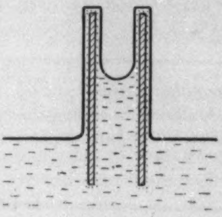


Figure 12.

Thus we can visualize the equilibrium of the system in the way indicated by figure 12, where, as before, the heavy lines represent surfaces in a state of tension γ_L and the dots represent the surface layer which is exerting a pressure $-\gamma_{L(L)}$. It will be seen from the diagram that, although the pressure $-\gamma_{L(S)}$ provides the initial raising force, it is not concerned in the final equilibrium which can be regarded as existing between the tension γ_L and the weight of the raised column. The film covering the walls of the tube above the meniscus is exactly equivalent, physically, to a soap film contained in a vertical frame dipping into a dish of liquid.

In practice, when a capillary tube is placed in a liquid the state of affairs depicted in figure 12 is, no doubt, never realized, because even when θ is zero the spontaneous spreading of a liquid over a solid surface does not, as a rule, take place. Nevertheless there will be some creeping of the liquid up the walls of the tube above the column—indeed, the method of ensuring zero contact angle is to wet the inside of the tube above the ultimate equilibrium position of the meniscus. The film thus formed may terminate at some place on the walls where spreading is arrested by an irregularity or variation in the nature of the solid surface but, for equilibrium, the film must nevertheless exert a downward force on the edge of the meniscus equal to $-\gamma_{L(S)}-\gamma_L$ per unit length. In many of the standard capillary-rise determinations, the tube and reservoir are contained in a closed vessel. In these circumstances if the walls of the tube were initailly wetted the likelihood of a stable film persisting indefinitely is much greater than when the tube is exposed to the air. The condition of the film would be influenced by two factors in addition to those already mentioned. In the higher parts of the tube it must be in equilibrium with a slightly lower vapour pressure, and also the net pressure ϕ which it exerts would be less on account of the variation of hydrostatic pressure in a vertical direction within the film itself.

It is evident that even if some mechanical energy is made available when the solid surface is wetted, this fact does not affect the above discussion of the equilibrium when the contact angle is zero. If the angle is not zero, however, we do not envisage a liquid film covering the walls of the tube above the meniscus, and any difference which may exist between the surface energies of the dry and wet solid surfaces must be written down as an upward force $\gamma_S-\gamma_{S(L)}$ acting on unit length of the edge of the meniscus in addition to $-\gamma_{L(S)}$. The relation between γ_L and h is, of course, unaltered by this modification.

§ 7. ACKNOWLEDGMENTS

The author wishes to thank Professor E. N. da C. Andrade, F.R.S., and Mr. D. O. Wood for reading the first draft of this lecture and for making valuable suggestions with regard to the method of presentation.

DISCUSSION

Dr. L. Hartshorn. I am surprised to hear that so many reputable physicists have had the temerity to assert that surface tension is fiction and surface energy real. Personally I doubt whether the word " real " has any place in physics, which I take to be the science of measurable properties. If by " real " you mean measurable, and by "fiction " not measurable, then you must admit that the surface tension of liquids is real because it is directly measurable, while surface energy is much more doubtful. Energy is of such dominating importance in unifying the various branches of physics, that we are apt to forget that it is not directly measurable, and to assume that its high status in theory necessarily gives it the same status in practice. A striking example occurs in electricity. Any elementary student, given a good potentiometer, can measure voltage with ease to an accuracy of say 0·01 per cent or better. This quantity is usually stated to be defined in terms of the energy expended in moving an electric charge from one place to another, a quantity which even the most accomplished experimentalist could not measure with an accuracy of 1 per cent. Is this good physics ? Which is the real quantity ? As we are discussing surface tension and not electricity I will only mention that Dr. Norman Campbell and I struck this oddity when making a critical examination of electricity, which we hope will appear in the *Proceedings*. Returning to surface tension, I hold that those who wish to make such distinctions may well say that in liquids surface tension is more real than surface energy, because it is more directly measurable : in solids, however, the tension is no longer directly measurable ; the two conceptions are on much the same footing, and surface energy may be preferred because energy is a more powerful conception in theory. I welcome the spirit of Dr. Brown's lecture in refusing to accept with the customary complacency traditional notions that are not satisfying: the interest it has aroused is a most encouraging sign.

Mr. C. GURNEY. I could not more agree with Dr. Brown in his criticism of the idea that the tension in surface of a liquid is a mathematical fiction. I believe it is as real as the tension in a piece of stretched wire. I would like to suggest, however, a rather different explanation of this tension.

The environment of the atoms in the surface of a liquid is different from that of those in the interior. The surface atoms have fewer nearest neighbours and consequently they are less strongly held to one another. It is therefore easier for thermal motions to eject atoms from the surface than from the interior; in other words, the tendency for atoms to escape from the surface is greater than the tendency for atoms to escape from the interior. This tendency of atoms to escape is measured by the chemical potential. The chemical potential of surface atoms is therefore higher than that of interior atoms. In equilibrium the chemical potential of the surface atoms must equal that of the interior atoms, and of course also that of the atoms in the vapour phase. How can this be achieved? It may be convenient here to digress and consider other cases of thermal equilibrium. Consider a pure substance, say lead. At atmospheric pressure the melting-point of lead is above room temperature. The chemical potential of solid lead is therefore less than that of liquid lead, so that the solid phase is stable. How could we have solid and liquid lead in equilibrium at room temperature. The answer is, that a suitable stress system must be applied, either to the solid or liquid or both. If we apply a hydrostatic pressure to the solid we could increase the chemical potential of the solid until it equalled that of the stress-free liquid. Two-phase equilibrium would then be possible. Or we could apply a hydrostatic tension to the liquid. This would decrease the chemical potential of the liquid until it equalled that of the stress-free solid. This gives the clue to the question of surface-liquid equilibrium. We can reduce the chemical potential of the surface by applying a tension to it. Tension reduces the tendency of atoms to escape from the surface, and at any given temperature there will be one value of the tension which reduces the tendency of atoms to escape from the surface so that it is equal to the tendency of atoms to escape from the liquid. Liquid and surface then have the same chemical potential and equilibrium is attained. For a liquid under hydrostatic pressure $\partial\mu/\partial p = \bar{V}$, where μ is chemical potential, p is pressure, V is specific volume. For a surface under tension $\partial\mu/\partial\gamma = -\bar{A}$, where γ is tension in surface and $\bar{A}$ is specific area.

A freshly formed surface having a given perimeter might have no tension in it. Atoms therefore escape from the surface faster than they enter it, until the spacing of atoms in the surface is further apart than the stress-free spacing. A tension is thus set up in the surface. It is not a dynamic tension as is thought to occur in rubber. It is just the same sort of tension as in a stretched wire, only unlike the tension in a stretched wire, the tension in the surface can never relax by creep. As soon as it tended to do so, more atoms would escape from the surface than would enter it, and the surface would be tightened up. There is no need to assume that surface atoms have any special rigidity or orientation. The liquid must have a surface and the surface cannot be stable unless it has a tension acting in its plane. Here, in a mobile liquid, is a tension that cannot relax. In an unstrained liquid the surface molecules leave the surface and enter the interior until a tension can be set up and equilibrated by a pressure within the liquid. The surface thus becomes spherical. The fact that the liquid in a drop is under pressure, causes the chemical potential of the liquid to be higher than that under a flat surface. A smaller tension in the surface is then sufficient to reduce the chemical potential of the surface to equal that of the liquid. Surface tension thus decreases as the radius becomes smaller. By equating the partial derivatives of the chemical potential of the liquid with respect to pressure to that of the surface with respect to tension we obtain a formula for the change of surface tension with radius. It is $\partial\gamma/\partial r = (2\gamma/r^2)(\bar{V}/\bar{A})$, where γ=surface tension, r=radius, $\bar{V}$=specific volume of liquid and $\bar{A}$=specific area of surface. $\bar{V}/\bar{A}$ is thus the effective thickness of the surface film. This equation does not seem to be generally known, but it is really a particular case of a more general equation given by Gibbs in 1876 (*Collected Works*, Longmans, New York, 1928, p. 232).

I would now like to refer to the surface tension of solids. If the solid is slowly cooled from the liquid state without phase change—for example, freshly made glass—it might have a tension in the surface approximating to the equilibrium value. Owing to viscosity, it would probably never attain the true equilibrium value, but it would tend to it for very slow cooling. If, on the other hand, the surface is formed by fracture, it would have no tension (except the polarization tension) in the surface to begin with, and would achieve its equilibrium tension so slowly that appreciable tension might never be developed. In this case the chemical potential of the surface layer would be higher than that of the underlying material, and the surface layer would be more chemically reactive.

In conclusion I would like to mention the tension in the surfaces of solids caused by polarization of ions and analogous effects. Surface atoms have fewer nearest neighbours, and consequently the bond strength between a surface atom and each of its remaining neighbours is greater than that between an interior atom and each of its greater number of neighbours. For example, in a covalent bonded material such as diamond, the bonds which are broken when the surface is formed, become resonating double bonds between the surface atoms and their remaining neighbours. The strength of a carbon double bond is greater than that of a single bond and the equilibrium atomic spacing is smaller. The surface atoms would therefore tend to move closer together, but they are prevented from doing so by the rigidity of the underlying solid. A tension is therefore set up in the surface. In an ionic bonded material such as sodium chloride, the surface ions become polarized, the concentration of electric charges becoming increased in the directions of the neighbouring ions and again the bond strength is increased. It is important to realize that the polarization tension is not the equilibrium surface tension, although it will reduce the chemical potential of the surface atoms and will thus reduce the additional tension which must be applied to achieve equilibrium.

Author's reply. It is pleasing to hear that both Dr. Hartshorn and Mr. Gurney agree with me in according a physical reality to surface tension. I am inclined to believe that those who profess to regard it as a fiction are not always really convinced by their own contention.

In his opening remarks Dr. Hartshorn implies that " reality " is a rather ill-defined concept and I entirely agree. In this lecture I have not been concerned with the philosophical question of the relative reality of force and energy in general. My sole purpose has been, so to speak, to ask those who concede a physical existence to surface energy, but not surface tension, whether they have any more reason for doing this than for regarding, say, the weight of a body as a fiction, but its gravitational potential energy as a reality. In other words, I believe that surface tension and surface energy are no more or less real than any other types of force and energy.

Dr. Hartshorn seems inclined to estimate the reality of physical quantities by the directness with which their values can be determined in particular cases. If the application of this criterion would make velocity less real than the length and time from which it is derived (except when it is read directly from the speedometer of a car !) I should be inclined to question its suitability.

Personally I doubt whether much is to be gained by trying to ascribe degrees of reality to force and energy in liquid and solid surfaces as Dr. Hartshorn suggests later on in his remarks. It seems to me that force is force and energy is energy wherever they are found and whatever may be their physical " cause ".

Mr. Gurney's remarks present the mechanism of surface tension in an interesting light. In his equation for the variation of surface tension with drop radius I wonder whether it is not necessary also to take into account the variation of vapour pressure with curvature. The possibility mentioned by Mr. Gurney of a difference of condition between solid surfaces, formed on the one hand by fusion and subsequent solidification and on the other by fracture, suggests that measurements of contact angle between a given liquid and the two types of surface might yield interesting results.

ULTRA-VIOLET ABSORPTION BAND-SYSTEMS OF PbO, PbS, PbSe AND PbTe

By E. E. VAGO and R. F. BARROW,
Physical Chemistry Laboratory, University of Oxford

MS. received 22 *October* 1946

ABSTRACT. A study has been made of the absorption spectra of PbO, PbS, PbSe and PbTe in the ultra-violet region ($\lambda > 2050$ A.). The vibrational analysis of new band-systems of PbS, PbSe and PbTe are reported, together with some observations on systems of PbO ($\nu_e \sim 35000$) and of PbSe ($\nu_e \sim 24818$ cm.$^{-1}$) already known. The table below includes the values of ν_e, ω_e and $x_e\omega_e$ at present available for the upper states of these molecules which are involved in absorption systems in this region of the spectrum.

State	PbO	PbS	PbSe	PbTe
G	—	—	—	46541·7
				159·6
				1·4
F	—	47770	45220·9	41658·8
		370	224·8	176·4
		(7·8)	0·50	1·0
E	34900	? 34000	—	—
	430	—		
	—	—		
D	30198·7	29650·5	28418·0	27176·5
	530·5	299·3	190·4	142·6
	2·92	1·57	0·53*	1·58

* $y_e\omega_e = -0.004$.

§ 1. INTRODUCTION

THE electronic states of the group of molecules comprising the oxides, sulphides, selenides and tellurides of C, Si, Ge, Sn and Pb were briefly reviewed a little while ago (Barrow, 1944): one conclusion was that several hitherto undetected states should exist, some of which should be involved in absorption band-systems lying in the near infra-red and in the ultra-violet regions of the spectrum. The ultra-violet region has since proved to be surprisingly fruitful, and we have recently described systems of the silicon and tin compounds occurring there (Vago and Barrow, 1946 a, b). In the present paper we give the results of a study of the absorption spectra of the lead compounds in the same region.

The arbitrary low-wave-length limit of about 2050 A. set by our present experimental arrangements implies that the work is far from complete. For example, we have found levels of PbTe at 41660 (F) and 46540 (G) cm.$^{-1}$: since, for corresponding systems, ν_e very often shifts to higher values with decreasing atomic number, we are led to predict an F level of PbO and G levels of PbO, PbS and PbSe which should take part in absorption systems in the region 50 000 to 65 000 cm.$^{-1}$ There is the same kind of assurance that SiO, GeO, SnO, SnS and SnSe will also absorb here, and somewhat less certain predictions may be made about many of the other molecules of this group. It is hardly possible to give an adequate discussion of the upper states of these molecules when so much experimental information is still lacking: for this reason we shall be concerned here primarily with the observations relating to the region 2050 to 3800 A.

§ 2. RESULTS

PbO

The absorption spectrum of PbO was studied in detail by Howell (1936), who recognized five band-systems and suggested the existence of a sixth. Of these, the system of shortest wave-length (called $F \leftarrow X$ by Howell, here E←X; see note to table 9) contained only seven bands. We therefore thought it of interest to examine this system again in the hope of extending it: at the same time we looked for absorption at somewhat shorter wave-lengths. We found no new bands in the range 2400–2700 A., but did obtain spectrograms of the E←X system which have led us to a rather different interpretation of this system.

The spectra were photographed in a large quartz Littrow instrument (dispersion about 3·3 A./mm. at 2800 A.) on Ilford Process plates. The effective path-length was estimated at 15 cm. and the temperature at about 1300° C. Bands of the D–X system of PbO are well developed under these conditions (see reproduction, figure 1) and at the short-λ end it is overlapped by the E←X system. Rotational structure lines, presumably of bands of the E←X system, can be traced to about 2550 A., but the strongest heads lie in the range 2800 to 3075 A.

Our band-head measurements are summarized in the Deslandres scheme in table 1. The arrangement there differs from that given by Howell by the inclusion

Table 1. Measurements of band heads of of the E←X system of PbO

Figures in parentheses are visual relative intensities.

	0		1		2		
3	35959 (1)* $2780\cdot1_8$	$715\cdot_5$	35243·5 (1)* 2836·57	701·9	34541·6 (1)* 2894·21		
	363		364·0		371·0		
2	35596 (1)* 2808·5	$716\cdot_8$	34879·5 (5)* 2866·17	708·9	34170·6 (3)* 2925·64		
	414		409·4		405·0		
1	35182 (0) 2841·5	712	34470·1 (4) 2900·21	704·5	33765·6 (3) 2960·73	699·6	33066·0 (2)* 3023·38
			422·3		425·5		424·2
0 ↑ v'			34047·8 (2) 2936·19	707·7	33340·1 (4) 2998·52	698·3	32641·8 (4) 3062·67
	$v'' \rightarrow 0$		1		2		3
$\Delta G''$ (mean):		714·7		705·8		699·0	
$\Delta G''$ (calc.):		714·4		707·0		699·6	

* Band also measured by Howell (1936).

PROC. PHYS. SOC. VOL. 59, PT. 3 (E. E. VAGO AND R. F. BARROW)

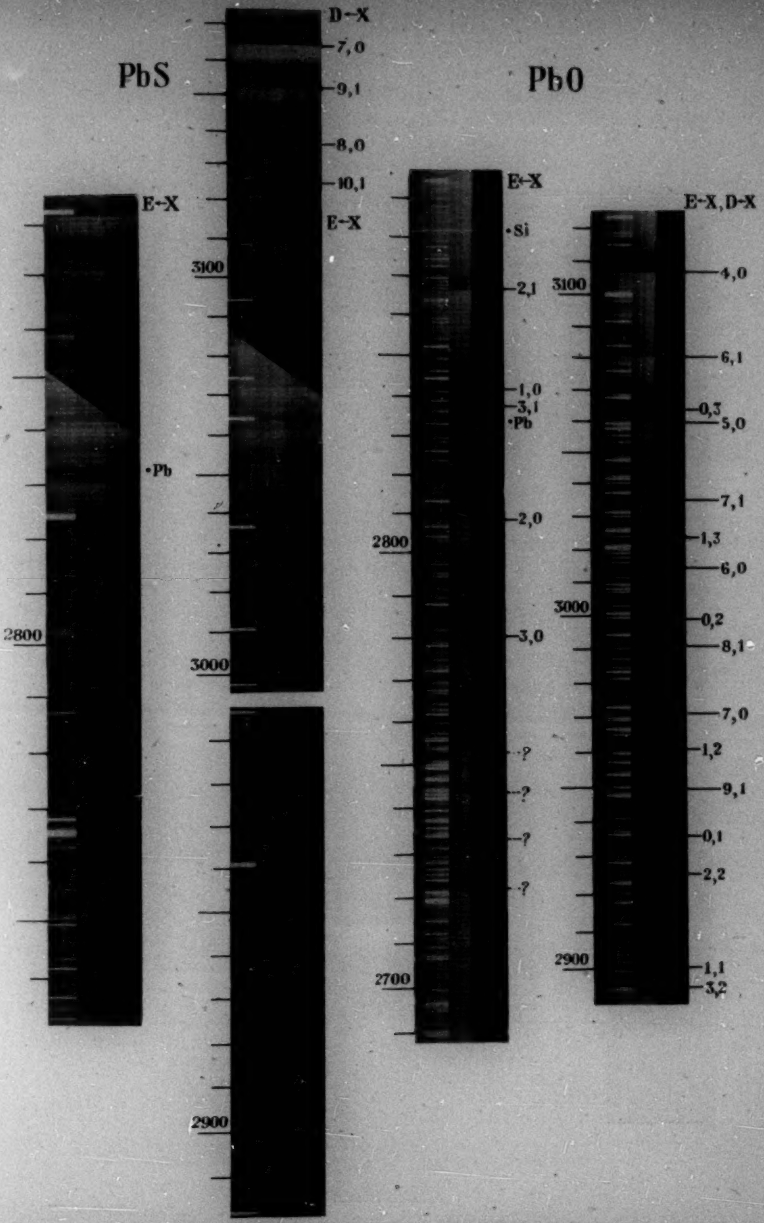


Figure 1. Bands of the D→X and E←X systems of PbO and PbS : large quartz Littrow spectrograph. (Fe and Cu arc reference lines.)

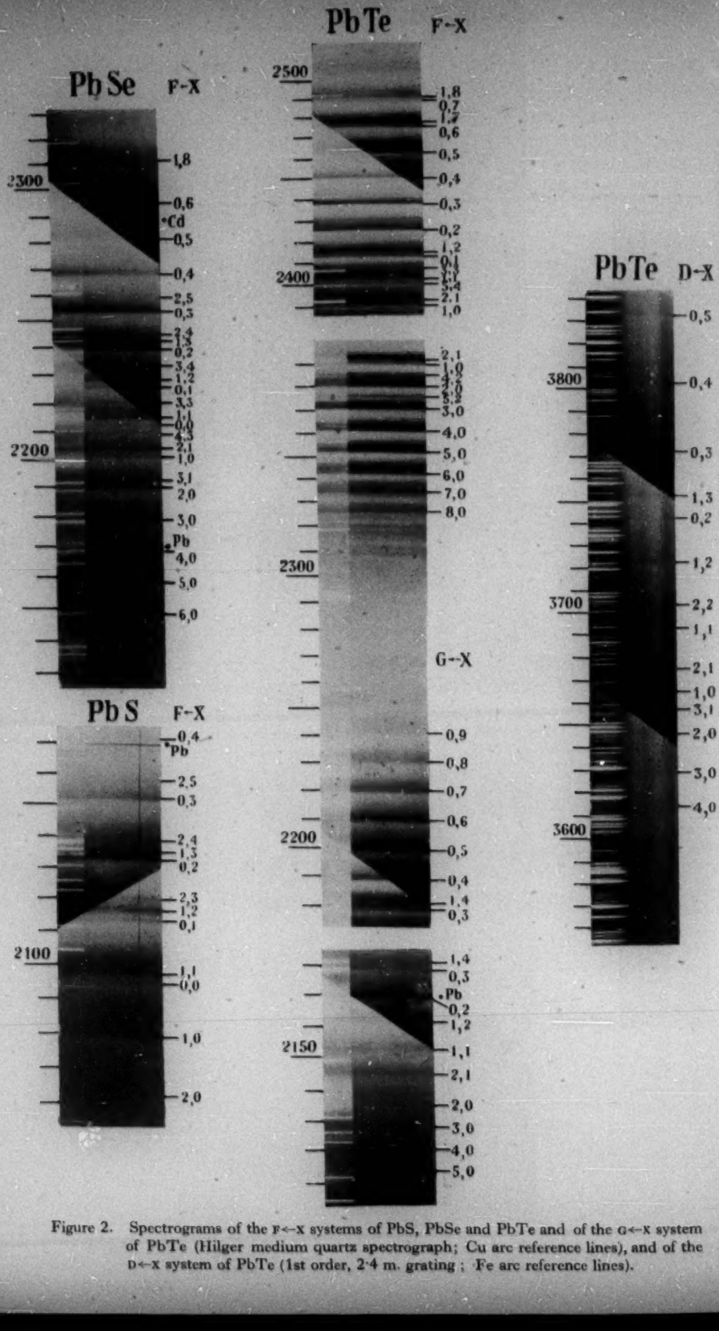


Figure 2. Spectrograms of the F←X systems of PbS, PbSe and PbTe and of the G←X system of PbTe (Hilger medium quartz spectrograph; Cu arc reference lines), and of the D←X system of PbTe (1st order, 2·4 m. grating; Fe arc reference lines).

of the bands with $v'=0$ and 1 (the 1,3 band was also measured by Howell but listed as unclassified). The two sets of measurements on the bands with $v'=2$ and 3 are reasonably concordant, but the intensities differ qualitatively. Thus for the two bands at 2780 and 2866 we find (roughly) 1 and 5, while Howell quotes 8 and 8: again we find very little evidence for a head at 2753 listed by Howell as occurring with intensity 6. These discrepancies suggested that some of the bands might be due to impurities, for it seems difficult to account for such variations in intensity on the basis of any plausible differences in temperature or in plate-sensitivity. We have not, however, been able to attribute any of the bands to another molecule.

It seems probable therefore that the new bands with $v'=0$ and 1 do belong to the E←X system. This view, and the correctness of the analysis given in table 1, are to some extent confirmed by the $\Delta G''_v$ values derived therefrom, which are seen to be very close to those calculated from Howell's expression determined from measurements of the other, more extensive, systems of PbO. The $\Delta G'_v$ values decrease irregularly and unusually rapidly as v' increases, suggesting strong perturbation of the E level.

PbS

The absorption spectrum of PbS has been investigated by Rochester and Howell (1935): their measurements extend to about 3135 A.—the short-λ end of the system with $\nu_e=29650$ (D–X, see note to table 9). We have now examined, and found bands at, both extremes of the region 2060–3200 A. Ilford Q1 plates and a Hilger medium quartz prism instrument (E.498: dispersion about 5·5 A./mm. at 2150 A.) were used for the range 2060–2200 A.: the range 2500–3100 A. was photographed in a first order of a 2·4-m. concave grating (linear dispersion about 7·4 A./mm.) and in the large quartz Littrow spectrograph on Ilford Process plates.

We have not been able to analyse the bands lying between 2750 and 3100 A. (table 2). The fact that many of them can be fitted into short progressions which yield differences (415–430 cm.$^{-1}$) of the order of the $\Delta G''_v$ values for PbS strongly suggests that they are indeed to be attributed to this molecule, but no likely

Table 2. Measurements of band-heads observed in the absorption spectrum of PbS

λ, I.A.	ν, cm.$^{-1}$	λ	ν	λ	ν
2754·85	36389	2852·0	35053	2974·35	33611*
57·8	350	63·7	34910*	76·8	583*
64·05	168	69·6	838*	86·8	471*
70·1	089*	70·9	822	97·15	355*
77·4	35994*	81·6	693*	3013·1	179*
84·65	901	92·25	565	24·15	057*
86·15	881	2904·6	418*	41·1	32873
93·95	781	15·7	287*	52·0	756*
2802·65	670*	30·6	113	62·8	640*
10·3	573*	41·15	33990*	71·8	545
21·25	435	52·15	863*	83·0	427
30·15	323*	59·25	782*	90·25	350*
41·55	182	64·65	721	3191·05	238*

* Indicates that the band may be fitted to a short v''-progression.

vibrational scheme has suggested itself; nor does it seem probable that all the bands can arise from high-v' transitions in the D–X system (which appears strongly in the plates). We are inclined to believe that some at least of the bands form an E←X system of PbS similar to that of PbO just described. At the short-λ end the bands crowd very closely together and have not been observed below about 2715 A., which may give a low-energy limit (~4·55 ev.) for a dissociation process.

The measurements and vibrational analysis of the bands in the region 2060–2200 A. (see figure 1) are given in table 3. The system (F←X) has not been observed in its entirety, as we were unable to photograph below about 2060 A., but enough has been measured—some twenty bands—to make the vibrational analysis reasonably certain. The v''-assignments follow from the known values of $\Delta G''_v$ (taken from Rochester and Howell), and there is then little doubt about the values of v' as the system is bounded by an intense progression with $v' = 0$. The upper-state vibrational levels appear to be rather irregular: the expression given in table 9 has been derived on the assumption that the heads of bands with $v' = 0$ are displaced so as to lie about 12 cm^{-1} too low. This expression fits the values for the other band-heads satisfactorily, but it may be criticized on the ground that the value of $x_e\omega_e$ is abnormally large (7·8). It is possible therefore that the perturbations extend to some of the other vibrational levels of the F state: this question could probably be elucidated by extending the observations to shorter wave-lengths.

Table 3. Measurements of band-heads of the F←X system of PbS

4					48215 *2073·4*	422	47792·7 *2091·70*					
					312		311					
3	§		48325 *2068·6*	422	47903 *2086·9*	421	47482 *2105·4*					46228 *2162·5*
			320				323					
2	48430 *2064·2*	425	48005 *2086·8*		—		47158·8 *2119·82*	412·3	46746·5 *2138·53*	418	46329 *2157·8*	
	338		338				331·5		341			
1	48092 *2078·7*	426	47666·5 *2097·24*	421·4	47245·1 *2115·95*	417·8	46827·3 *2134·83*	421	46406 *2154·2*			
	363		362·4		366·9		371·2		367			
0 ↑ v'	47729·5 *2094·47*	425·4	47304·1 *2113·31*	425·9	46878·2 *2132·51*	422·1	46456·1 *2151·89*	417	46039 *2171·4*			
$v'' \to$	0		1		2		3		4		5	6

§ The limit of the observations was about 2060 A.

PbSe

One ultra-violet system of PbSe (D–X) has been described (Barrow and Vago, 1944). This was first obtained in emission, but spectrograms have since been taken in absorption, using the Hilger medium quartz instrument. Measurements of the new plates have confirmed the original vibrational analysis and have enabled several additional bands to be assigned (table 4). The expression for G'_v has been re-calculated and the result is given in table 9. Bands can be traced on the plates down to about 2870 A., which gives a minimum value for a dissociation limit at 4·3 ev. above the ground state.

Absorption bands have also been found in the region 2150–2300 A. (figure 2). These form a single system, F←X, whose vibrational analysis is given in table 5.

Table 4. Additional or amended measurements of bands of the D—X system of PbSe

v', v''	ν, cm.$^{-1}$	v', v''	ν, cm.$^{-1}$
14,1	30637	8,1	29585·5*
13,1	468	6,1	214·4*
14,2	364	4,0	121·6*
12,1	297	2,0	28751·9*
13,2	193	3,1	658·8*
11,1	120	0,0	373·8*
12,2	016	1,1	287·2*
10,1	29944	2,2	202·5*
11,2	845	0,1	098·2*
12,3	749	1,2	011·6*

* Band also observed in emission.

Table 5. Measurements of band-heads of the F←X system of PbSe

5	46309 *2158·7$_5$*				45747 *2185·2$_8$*									
	219													
4	46090 *2169·0*	277	45813 *2182·1*				45257 *2208·9*				44711 *2235·9*			
	220		221				224							
3	45870 *2179·4*	278	45592 *2192·7*				45033 *2219·9*	274	44759 *2233·5*					
	223		225						220					
2	45647 *2190·0$_8$*	280	45366·9 *2203·56*	278·0	45088·9 *2217·15*				44539 *2244·5*	270	44269 *2258·2*			
	228		225·1		221·8									
1	45419·3 *2201·02*	277·5	45141·8 *2214·55*	274·7	44867·1 *2228·11*	274·8	44592·3 *2241·84*							43503 *2298·0*
	224·3		223·7		224·8		223·4							
0 ↑ v'	45195·0 *2211·94*	276·9	44918·1 *2225·58*	275·8	44642·3 *2239·33*	273·4	44368·9 *2253·13*	273·6	44095·3 *2267·11*	273·6	43821·7 *2281·27*	269·0	43552·7 *2295·36*	
$v''\rightarrow$	0		1		2		3		4		5		6	7

PbTe

Several unsuccessful attempts were made to excite a D–X system in emission, using methods which had proved effective for SnSe and PbSe. We then examined the absorption spectrum and found not only the expected D←X system, but two other systems, F←X and G←X, as well.

The D←X system lies in the region 3550–3900 A. (figure 2): it was photographed in a first order of the grating instrument. The vibrational analysis is straightforward and is given in table 6.

Table 6. Measurements of band-heads of the D←X system of PbTe

v' ↑ \ v'' →	0		1		2		3		4		5		6
6	27933 *3579·0*												
	125												
5	27808 *3595·1*												
	129												
4	27679 *3611·8*												
	131												
3	27548·0 *3628·99*	210·9	27337·1 *3657·00*										
	130·2		131·1										
2	27417·8 *3646·23*	211·8	27206·0 *3674·62*	211·7	26994·3 *3703·44*	209·0	26785·3 *3732·33*						
	137·4		137·4		137·5		140·1						
1	27280·4 *3664·60*	211·8	27068·6 *3693·27*	211·8	26856·8 *3722·39*	211·6	26645·2 *3751·96*	212·0	26433·2 *3782·05*				
	137·7		138·2		138·0		135·7		134·3				
0	27142·7 *3683·19*	212·3	26930·4 *3712·23*	211·6	26718·8 *3741·62*	209·3	26509·5 *3771·16*	210·6	26298·9 *3801·37*	212·6	26086·3 *3832·35*	212·2	25874·1 *3863·78*

The F←X and G←X systems were photographed in the medium quartz prism spectrograph. The bands of the F←X system (figure 2) lie between 2300 and 2500 A. : again there is little doubt about the vibrational analysis (table 7), but the head of the 6,0 band appears to be displaced about 12 cm^{-1} from its expected position.

The strongest bands of the G←X system (figure 2, table 8) form a $v'=0$ progression in the region 2160–2250 A. Altogether seventeen bands have been assigned to this system: it is likely that more could be observed at shorter wavelengths, which would make possible the derivation of a more accurate expression for G'_v than that given in table 9, but it does not seem probable that the analysis would be changed thereby.

§ 3. SUMMARY

The results are summarized in table 9. It is not proposed to discuss them in detail, but it is perhaps of interest to compare them with those for the corresponding tin compounds. This is done in table 10, which includes the values of ν_e and of the force-constants expressed as percentages of those in the ground states. There seems to be some evidence that the latter values are grouped significantly, i.e. that the states labelled with similar letters are indeed spectroscopically similar, but perhaps the most useful function of this kind of table at the moment is to demonstrate the gaps in our knowledge. However, all the spaces in table 10 do not necessarily imply the existence of unknown stable states: thus, it may be that E←X systems of PbSe and PbTe do not exist, for the dissociation energies of the E states of the tin compounds are probably rather small, decreasing from oxide to telluride (Vago and Barrow, 1946 b), so that the corresponding levels of PbSe and PbTe may be unstable. This argument does not apply to the other states in

Table 7. Measurements of band-heads of the F←X system of PbTe

v′↑ v″→	0		1		2		3		4		5		6	
12					43181 *2315·1*									
					156									
11			43236 $2312{\cdot}1_8$	211	43025 *2323·5*									
			155											
10	43293 $2309{\cdot}1_8$	212	43081 *2320·5*											
	159		160											
9	43134 $2317{\cdot}6_8$	213	42921 $2329{\cdot}1_8$											
	159		158											
8	42975 *2326·2*	212	42763 $2337{\cdot}7_8$											
	156													
7	42819 *2334·7*													
	150													
6	42669·4 *2342·88*													
	172·8													
5	42496·6 *2352·41*				42067·2 *2376·42*				41643·9 *2400·58*					
	168·1				166·9									
4	42328·5 *2361·75*				41900·3 *2385·89*	211·1	41689·2 *2397·97*				41270·0 *2422·33*			
	169·9						168·8							
3	42158·6 *2371·27*	212·1	41946·5 *2383·26*				41520·4 *2407·72*						40889·6 *2444·87*	
	169·4		169·6											
2	41989·2 *2380·84*	212·3	41776·9 *2392·94*						41142·1 *2429·86*	210·0	40932·1 *2442·33*			
	172·0		172·1								165·6			
1	41817·2 *2390·63*	212·4	41604·8 *2402·83*	212·2	41392·6 *2415·15*	209·3	41183·3 *2427·43*				40766·5 *2452·25*	210·5	40556·0 *2464·98*	211·1
			177·5		175·7		177·5				179·9		177·5	
0			41427·3 *2413·13*	210·4	41216·9 *2425·45*	211·1	41005·8 *2437·94*	209·5	40796·3 *2450·46*	209·7	40586·6 *2463·12*	208·1	40378·5 *2475·82*	207·7

v′↑ v″→	7		8
1	40344·9 *2477·88*	211·5	40133·4 *2490·94*
	174·1		
0	40170·8 *2488·62*		

the table, however, and if the arrangement is correct, nine more F←X and G←X systems of these molecules await discovery.

Table 8. Measurements of band-heads of the G←X system of PbTe

4	47126 *2121·3*												
	151												
3	46975 *2128·1*	213	46762 *2137·8*										
	154		151										
2	46821· *2135·1*	210	46611 *2144·7$_8$*										
			155										
1	—		46456 *2151·9*	207	46248·7 *2161·54*	208·5	46040·2 *2171·33*	208	45832 *2181·2*				
					158		159·5		161				
0 ↑ v'	—		—		46091 *2168·9$_8$*	210	45880·7 *2178·88*	209·8	45670·9 *2188·89*	212·2	45458·7 *2199·11*	208·8	45249·9 *2209·26*	209
$v''\to$	0		1		2		3		4		5		6	

0 ↑ v'	45041 *2219·5*	210	44831 *2229·9*	212	44619 *2240·5*
$v''\to$	7		8		9

Table 9. Values of ν_e, ω_e and $x_e\omega_e$

State	PbO	PbS	PbSe	PbTe
G	—	—	—	46541·7§ 159·6 1·4
F	—	47770 370 (7·8)	45220·9§ 224·8 0·50	41658·8§ 176·4 1·0
E	34900‡ 430 —	?34000 — —	—	—
D	30198·7 530·5 2·92	29650·5‡ 299·3 1·57	28418·0 190·4 0·53*	27176·5 142·6 1·58
X	0 721·8 3·70	0 428·1 1·20	0 277·4 0·56	0 211·8 0·12

* $y_e\omega_e = -0{\cdot}004$.

‡ It has seemed best to give the same letters to states which appear to be analogous. Unfortunately, this has led to two alterations. For PbO, the level with $\nu_e \sim 35000$ was formerly F: it is doubtful, however, if the level formerly known as E really exists (Howell, 1936), and we prefer this letter for the 35000 state as it emphasizes some similarities with the E levels of SnO, etc. The level of PbS at 29650 (formerly called E) is probably analogous to what are already known as the D levels of PbO and of PbSe, and it is has, therefore, been called D above.

§ A brief notice of the results of vibrational analysis of these systems has recently been published by Sharma (1946). His constants for the F←X systems of PbSe and PbTe agree quite closely with ours, but he obtains $\nu_e = 45918$ for the G←X system of PbTe.

Table 10. Values of ν_e (in ev.) and of the force-constants expressed as percentages of those in the ground states

State	SnO	PbO	SnS	PbS	SnSe	PbSe	SnTe	PbTe
G	—	—	—	—	—	—	5·86 78	5·77 56·8
F	—	—	—	5·92 75	5·93 77	5·61 65·7	5·46 78·4	5·16 69·4
E	4·56 36·6	4·33 35·5	4·10 36·4	? 4·2 —	3·81 35·1	— —	$3{\cdot}4_7$? 33	— —
D	3·68 50·2	3·74 54·0	3·51 46·3	3·68 48·9	3·42 46·2	3·52 47·1	3·15 47·6	3·37 45·3

REFERENCES

Barrow, R. F., 1944. *Proc. Phys. Soc.*, **56**, 204.
Barrow, R. F. and Vago, E. E., 1944. *Proc. Phys. Soc.*, **56**, 78.
Howell, H. G., 1936. *Proc. Roy. Soc.*, A, **153**, 683.
Rochester, G. D. and Howell, H. G., 1935. *Proc. Roy. Soc.*, A, **148**, 157.
Sharma, D., 1946. *Nature, Lond.*, **157**, 663.
Vago, E. E. and Barrow, R. F., 1946 a. *Proc. Phys. Soc.*, **58**, 538 ; 1946 b. *Ibid.*, **58**, 707.

THE LANDAU VELOCITY IN LIQUID HELIUM II

By D. V. GOGATE and P. D. PATHAK,
Physics Department, Baroda College, Baroda, India

MS. received 11 *February* 1946 ; *in revised form* 21 *November* 1946

ABSTRACT. A simple derivation of Landau's expression for the second velocity of sound (in addition to the usual velocity $\sqrt{(dp/d\rho)}$) in helium II is given from general physical principles.

§ 1. INTRODUCTION

The peculiar properties of liquid He II have been the subject of considerable experimental and theoretical research during the last few years. At 2·19° K. (the so called λ-point) helium shows a discontinuity in its specific heat, indicating a characteristic type of phase transition. Its viscosity decreases suddenly at the λ-point and the entropy difference between the liquid and the solid phase tends towards zero, showing that the liquid phase goes into a peculiar state below the λ-point. As is well known, heat flow in liquid He II is accompanied by a transfer of momentum (the thermo-mechanical effect). It has further been found that different experimental arrangements in the measurement of viscosity and thermal conductivity lead to different values, e.g. the viscosity vanishes when measured by the flow of liquid helium through a thin capillary or a narrow slit, while a non-zero value of viscosity is obtained when the latter is measured by the rotating-disc method. A similar state of things is found to exist with regard to the values of thermal conductivity. The systematic and thorough experimental investigation of the peculiar properties of He II is due to P. L. Kapitza (1938). He has shown the thermodynamic reversibility of the thermomechanical effect and has thus evolved a method of attaining temperatures approaching the absolute zero.

The unusual characteristics of He II led L. Tisza (1938) and F. London (1938) to suggest that liquid He II may be regarded as an ideal degenerate Bose-Einstein gas. Tisza suggested that the atoms found in the normal state (a state of zero energy) move through the liquid without friction, while London tried to explain the discontinuity in the specific heat at the λ-point, and the thermomechanical effect, by treating helium II as a Bose-Einstein degenerate system in which one fraction of the substance is distributed over the excited states in a way determined by the temperature, while the rest is condensed in the lowest energy level.

Recently Mendelssohn (1945) has advanced the hypothesis that both the superfluidity of He II and the superconductivity of electrons are caused by one and the same mechanism of "frictionless transport". The superfluid helium atoms, like the superconducting electrons, form an aggregate in momentum space of zero thermal energy (*Z*-state). The *Z*-particles have zero entropy but an appreciable zero-point energy. This theory is not based on any particular theoretical conception but is built up on the existing experimental evidence.

However, the most noteworthy theoretical study of He II is due to L. Landau (1941). In a most suggestive paper he has not only accounted for some of the previously known peculiar properties of He II, but has also predicted the existence of two velocities of sound in that liquid. This prediction has been experimentally confirmed by V. Peshkov (1944), who has obtained a value of about 20 m./sec. for the second (Landau) velocity in He II.

Landau has introduced the concept of a quantum liquid of which He II provides, so far, the only practical illustration. From the quantization of an arbitrary system of interacting particles (a liquid) he has advanced a system of hydrodynamic equations describing the macroscopic motion of liquid He II and has shown that two velocities of sound must exist in that liquid at non-zero temperatures. As the existence of a second velocity (in addition to the usual velocity $\sqrt{(dp/d\rho)}$) in He II is rather interesting, and has been experimentally verified, it may be worth while to discuss the existence of this abnormal velocity (which will be referred to hereafter as the Landau velocity) from general physical principles. This is done in this paper.

§ 2. ELEMENTARY DERIVATION OF THE EXPRESSION FOR THE LANDAU VELOCITY

Following Landau, we assume that liquid He II contains two types of fluid. The first type (n) is the ordinary (normal) type of fluid, having for its viscosity η and entropy S their usual (non-zero) values. The second type of fluid (s) (which we may call a superfluid) possesses no viscosity ($\eta=0$) and no entropy ($S=0$). Let ρ_n, V_n and ρ_s, V_s denote the density and velocity of motion for the fluids n and s respectively.

Then the density ρ of liquid He II will be given by

$$\rho=\rho_n+\rho_s. \qquad \ldots\ldots(1)$$

The two types of fluid in He II do not resist the motion of each other, i.e. they possess independent motions. Thus for the flow of liquid He II we may write the equation

$$\rho V=\rho_n V_n+\rho_s V_s, \qquad \ldots\ldots(2)$$

where V denotes the average velocity of the fluids n and s.

We can contemplate a motion of the liquid when $V=0$. Then equation (2) will be reduced to

$$\rho_n V_n+\rho_s V_s=0, \qquad \ldots\ldots(3)$$

which means that there is no net transport of mass across any plane in the liquid.

Let us now consider the transport of entropy. For simplicity we consider the one-dimensional case. If we imagine a parallelopiped of unit cross-section and length Δx to be situated in the liquid with one pair of its opposite parallel faces vertical, the amount of entropy entering the parallelopiped per second at one face will be

$$\rho_n V_n S_n=\rho V_n S, \qquad \ldots\ldots(4)$$

where S is the entropy possessed by one gram of liquid He II.

The amount of entropy leaving the parallelopiped at the opposite parallel face per second will be

$$V_n\rho S+\frac{\partial}{\partial x}(V_n\rho S)\Delta x. \qquad \ldots\ldots(5)$$

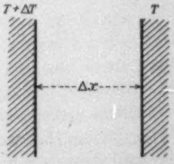
The net transport of entropy per unit volume per second is given by

$$\frac{\partial}{\partial x}(\rho S V_n)=-\frac{\partial}{\partial t}(S\rho). \qquad \ldots\ldots(6)$$

If we assume ρ to be sensibly constant and also take V_n to be small we obtain

$$\frac{\partial S}{\partial t}=-S\frac{\partial V_n}{\partial x}, \qquad \ldots\ldots(7)$$

neglecting the second order term $V_n\dfrac{\partial S}{\partial x}$.



Again the amount of entropy leaving the face at the temperature $T+\Delta T$ and reaching the opposite parallel face at temperature T will be given by

$$-\rho S V_n. \qquad \ldots\ldots(8)$$

Now we know from the second law of thermodynamics that $W/Q=\Delta T/T$, where W is the work done and Q the heat taken in at temperature T, or

$$\frac{\text{Work done}}{\text{Entropy absorbed by the system}}=\Delta T. \qquad \ldots\ldots(9)$$

We assume a reversible change in the case of liquid He II, and this is an essential condition for the existence of the abnormal velocity. Then, according to the second law of thermodynamics, the amount of entropy (which remains constant and which flows from one face to the other) multiplied by the difference of temperature between the two opposite parallel faces will be equal to the work done. Hence the entropy flow $\rho S V_n$ per second must produce work at the rate of $-(\rho S V_n)\Delta T$, and this will appear in the form of kinetic energy.

Now the kinetic energy per unit volume is given by $\frac{1}{2}[\rho_n V_n^2 + \rho_s V_s^2]$ and the rate of change of kinetic energy per unit volume is

$$\begin{aligned}
\rho_n V_n \frac{\partial V_n}{\partial t} + \rho_s V_s \frac{\partial V_s}{\partial t} &= \rho_n V_n \frac{\partial}{\partial t}(V_n - V_s) \text{ by equation (3)} \\
&= \rho_n V_n \frac{\partial}{\partial t}\left(V_n \frac{\rho}{\rho_s}\right) = \rho\rho_n V_n \frac{\partial}{\partial t}\left(\frac{V_n}{\rho_s}\right) \\
&= \rho\rho_n V_n \left[\frac{1}{\rho_s}\frac{\partial V_n}{\partial t} - \frac{V_n}{\rho_s^2}\frac{\partial \rho_s}{\partial t}\right] \\
&= \frac{\rho\rho_n V_n}{\rho_s}\frac{\partial V_n}{\partial t}, \qquad \text{......(10)}
\end{aligned}$$

neglecting the second-order term $\frac{V_n}{\rho_s^2}\frac{\partial \rho_s}{\partial t}$.

Hence the work done in volume Δx per second is

$$\frac{\rho\rho_n V_n}{\rho_s}\frac{\partial V_n}{\partial t}\Delta x, \qquad \text{......(11)}$$

and by the second law of thermodynamics this must equal $-(\Delta T)\rho S V_n$.

Hence we have

$$S\frac{\partial T}{\partial x} = -\frac{\rho_n}{\rho_s}\cdot\frac{\partial V_n}{\partial t}. \qquad \text{......(12)}$$

Again we know $dQ = c\,dT + p\,dv = T\,dS$, and when v is constant, $\partial T/\partial S = T/c$, where c is the specific heat.

Thus

$$\frac{\partial T}{\partial S} = \frac{T}{c} = \frac{\partial T}{\partial x}\frac{\partial x}{\partial S}$$

or

$$\frac{\partial T}{\partial x} = \frac{T}{c}\frac{\partial S}{\partial x}.$$

Hence from (12) we get

$$S\frac{T}{c}\cdot\frac{\partial S}{\partial x} = -\frac{\rho_n}{\rho_s}\cdot\frac{\partial V_n}{\partial T}$$

or

$$\frac{\partial V_n}{\partial T} = -\frac{\rho_s}{\rho_n}\cdot S\frac{T}{c}\cdot\frac{\partial S}{\partial x}. \qquad \text{......(13)}$$

Differentiating (7) with respect to t and (13) with respect to x, we get

$$\frac{\partial^2 S}{\partial t^2} = \frac{\rho_s}{\rho_n}\cdot\frac{S^2 T}{c}\cdot\frac{\partial^2 S}{\partial x^2} \qquad \text{......(14)}$$

or

$$u^2 = \frac{\rho_s}{\rho_n}\cdot\frac{S^2 T}{c} \qquad \text{......(15)}$$

where u (or u_2) is the second velocity (Landau velocity) of sound.

REFERENCES

Kapitza, P. L., 1938. *Nature, Lond.*, **141,** 74 ; 1940. *J. Phys. U.S.S.R.*, **4,** 181 ; 1941. *Phys. Rev.*, **60,** 354.

Landau, L., 1941. *Phys. Rev.*, **60,** 358 ; *J. Phys. U.S.S.R.*, **5,** 71 ; 1944. *J. Phys. U.S.S.R.*, **8,** 1.

London, F., 1938. *Nature, Lond.*, **141,** 643 ; 1939. *J. Phys. Chem.*, **43,** 49.

Mendelssohn, K., 1945. *Proc. Phys. Soc.*, **57,** 371.

Peshkov, V., 1944. *J. Phys., U.S.S.R.*, **8,** 381.

Tisza, L., 1938. *Nature, Lond.*, **141,** 913.

THE RADIO DETECTION OF METEOR TRAILS AND ALLIED PHENOMENA

By Sir EDWARD APPLETON, F.R.S. and R. NAISMITH

Read 31 *January* 1947; *MS. received* 21 *February* 1947

ABSTRACT. The results of a series of radio sounding observations on (*a*) transient bursts of atmospheric ionization, and (*b*) abnormal or sporadic E-layer ionization, are described and discussed. Special observations during the recent Giacobinid meteor shower of 10 October 1946 confirm that the ionization bursts which can be observed at all hours of the day are due to sporadic meteors. A statistical study of abnormal or sporadic E-layer ionization gives strong support to the view that, in temperate latitudes, this phenomenon is also largely ascribable to the same cause.

§1. INTRODUCTION

It has been known for many years that the ionization in the E-layer of the ionosphere is subject to sporadic changes. Such changes are due to the incidence of irregular ionizing agencies which are additional to the normal ultra-violet radiation from the sun. In 1935 (Appleton and Naismith, 1935), for example, evidence was presented which indicated the separate existence and unlike behaviour of abnormal and normal E-layer ionization and by which it was possible to demonstrate that the normal ionization due to solar ultra-violet light followed a very regular law as regards its dependence on solar zenith distance; and that the great variability of E-layer ionization as a whole was due mainly to the variability of the additional agency responsible for abnormal ionization.

In this paper we describe the results of a series of radio investigations, carried out in recent years, which shed some light on the origin of the irregular ionization in the lower ionosphere. Using the ordinary methods of radio sounding by pulses, two types of this sporadic ionization have been recognized. It may well be, as suggested below, that these have a common origin, but the two manifestations of it are sufficiently dissimilar in character for us to treat them separately.

We shall call the two types of ionization:—Type (A), E-layer bursts of ionization, and Type (B), Abnormal or sporadic E-layer ionization; and discuss them in the order given.

§2. E-LAYER IONIZATION BURSTS (TYPE A)

During the International Polar Year 1932–3 a transient type of echo was observed (Appleton, Naismith and Ingram, 1937) at about the level of the E-layer. This echo was found to last only a second or two, and occurred both by night and day. The fact that there was no major difference in incidence between night and day was considered to exclude solar ultra-violet light as the responsible agency; and it was pointed out that, possibly, meteors were responsible for these ionization bursts, since Skellet (1935) had observed that major increases of Abnormal E ionization (i.e. Type (B) above) occurred at night when meteors were observed to pass overhead.

A fairly extensive study of E-layer bursts of ionization was later made by Appleton and Piddington (1938), at Cambridge, who used a 2 to 3 kilowatt pulse sender on 8·8 Mc./s. coupled to a dipole aerial which radiated vertically upwards. These authors made a study of the distribution of echo occurrence in range and found that this distribution was very similar by day and by night. They also measured the average equivalent reflection coefficient of ionization bursts, for the particular frequency employed. The same type of observations was continued at Cambridge by Mohanty (1938), in an extensive series of experiments, who showed that the reflections from ionization bursts continued to be appreciable even when the frequency used was increased to 16 Mc./s., although the effective reflection coefficient was found to be less as the frequency was increased. Mohanty further came to the conclusion that the ionization bursts (Type A) were the trials of meteors.

§3. ABNORMAL OR SPORADIC E-LAYER IONIZATION (TYPE B)

Using radio frequencies slightly in excess of the critical frequency of the normal E-layer, it is usually possible to detect reflections from more intensely ionized clouds or strata embedded in the normal E-layer. Such clouds or strata are called *abnormal or sporadic E-layer ionization*. They differ from ionization bursts in that they are more persistent in time and, in daytime, appear at a very limited range of heights. They are particularly in evidence in summer and in the daytime. On occasion, and most frequently in summer, this ionization may reach such high densities that its critical penetration frequency exceeds that of the F_2-layer. We have referred to such manifestations as "Intense E-ionization" (Appleton, Naismith and Ingram, 1939; Appleton and Naismith, 1940). Under conditions of marked abnormal E-layer ionization, the upper limit of frequency usuable for long-distance communication may be abnormally high, in that the usual restriction by way of F_2-layer electron limitation is not operative.

Abnormal or sporadic E-layer ionization, at any rate in the daytime, is usually situated somewhere near the level of the maximum electron density of the normal E-layer. Reflections from it are also fairly markedly frequency-dependent in that, as the radio frequency employed is increased, the intensity of reflection falls off fairly rapidly within a definite frequency range. Before echoes become inappreciable there is, however, often a frequency interval of partial reflection in which echoes are simultaneously observed from both the abnormal E-layer and from the F-layer.

§4. OBJECT OF THE PRESENT SERIES OF EXPERIMENTS

The present series of observations was undertaken with the general object of comparing the incidence of Type A and Type B ionization in the E-layer and, in particular, of testing how far the meteor theory accounted for Type A.

For the experiments on ionization bursts, a high-power pulse sender and a receiver were kindly lent to us by the Air Ministry. In view of the fact that the earlier observations had been conducted on the lower range of frequencies it was decided to use a higher radio-frequency of 27 Mc./s. These observations were conducted on a flat site at the Radio Research Station where an aerial system designed to radiate vertically upwards was erected for us by 60 Group of the Royal Air Force. Pulses of 15 μs. and of a repetition rate of 50 per second were used and the usual cathode ray oscillograph display of ground and echo pulses was employed.

For the study of abnormal or sporadic E-layer ionization we were able to rely on results from the daily soundings of the ionosphere conducted at the Slough Radio Research Station, which form part of the continuous series of measurements of ionospheric layer densities which we have made since 1931, when the critical-frequency method was first introduced (Appleton, 1931; Appleton and Naismith, 1932).

For convenience we describe the observations on the two types of abnormal ionization in two sections below dealing respectively with Type A and Type B.

§5. EXPERIMENTAL RESULTS: PART I

In this section we deal with the observations on ionization bursts using a radio exploring frequency of 27 Mc./s. By employing photographic registration we have endeavoured to eliminate the subjective feature of the earlier observations made in Cambridge. We have also tried to keep the experimental conditions as constant as possible over a period of over two years, so that the seasonal variation of ionization bursts could be examined. Supplementary eye observations have, however, also been made. Special series of observations have also been made at times of expected meteor showers, notably on 10 October 1946, on the occasion when the Giacobini-Zinner comet approached the earth.

Perhaps the most remarkable result of the continuous series of observations is the seasonal variation of ionization-burst occurrence at noon. In figure 1 are shown the monthly means of the number of bursts per hour at noon for the period 1944–1946. It will be seen that, although summer values are higher than the winter values, the curve is not symmetrical about the summer solstice.

(In the same figure are shown, for comparison, the monthly means of the frequencies of occurrence of noon sporadic E-layer ionization on 4 Mc./s. This comparison will be referred to below.)

In our study of the diurnal variation of ionization bursts, attention has been specially directed to nocturnal events, so that a comparison could be made with

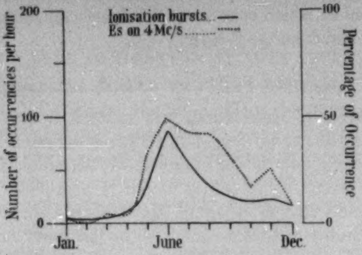


Figure 1. Seasonal variation of frequency of ionization bursts and of frequency of occurrence of sporadic E-layer ionization (4 Mc./s.) in daytime. (Average 1944–6.)

the results of visual meteor observations. In figure 2 is illustrated the nocturnal variation of ionization-burst frequency, from which it will be seen that the rate of ionization-burst detection is greater after midnight than before. In figure 3 is shown the nocturnal variation of ionization-burst duration, showing that the

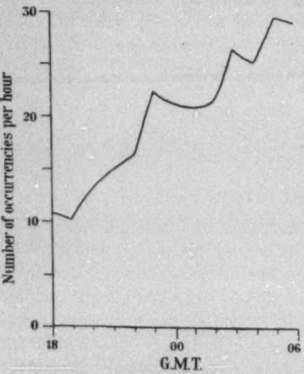


Figure 2. Nocturnal variation of ionization burst frequency.

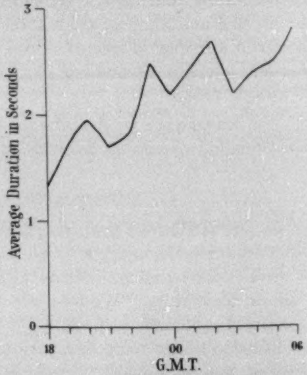
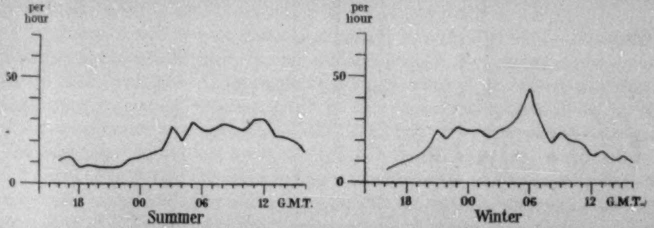


Figure 3. Nocturnal variation of ionization burst duration.

duration of these transient echoes is also greater after midnight than before. Since we know that, for constant experimental conditions, echo duration usually increases with echo intensity, we interpret the latter result as indicating that stronger reflections from ionization bursts are obtained in the second half of the night than in the first.

In figure 4 are shown typical diurnal curves of ionization-burst frequency in summer and winter. A remarkable difference is noted between these two curves. The night-time variation is found to be fairly constant in character, but the day-time portion alters with the season of the year. For example, it is found that there is a characteristic minimum about 1800 G.M.T. throughout the year. It is also found that the number of bursts increases from that time onwards until about 0500 G.M.T. in the winter, the average number of bursts in the second half of the night being about twice the value during the first half. Under these winter conditions the early morning maximum is the maximum of the day. On the other hand, in summer, the increase in echo occurrence in the second half of the night continues after dawn so that the maximum of the day is reached just before, or at, local noon.

Before attempting to interpret the above results in terms of the meteor theory of origin, we describe first what we think is the most weighty evidence for this theory of which we are aware. Ionospheric literature contains many examples of simultaneous experimental observations on abnormal E-layer ionization and visual observations on meteors, and many workers have concluded that such ionization was increased at such periods of meteoric activity (Skellet, 1931; Mitra, Syam and Ghosh, 1934; Bhar, 1937; Pierce, 1938). More recently it has been considered desirable to concentrate such radio observations, not on abnormal E-ionization but on ionization bursts, and Hey and Stewart (1946) have recently shown a correlation between ionization bursts and meteor showers. By making daily observations with a radio beam directed vertically upwards, they found marked peaks in the frequency of occurrence of ionization bursts coinciding with major meteor showers. Further, it was found that stations with inclined radio beams gave different diurnal variations when set on different bearings; by making the assumption that the most favourable direction of observation was at right angles to the meteor train, they were in certain cases able to derive from these results the approximate positions of the meteor radiants.



Summer

Winter

Figure 4. Illustrating difference in diurnal curves of ionization burst frequency (average number of occurrences per hour) in summer and winter.

We have found similar correlations between increases of ionization-burst frequency and the incidence of meteor showers, but not one of them is nearly as striking as that which occurred on 10 October 1946 on the occasion of the shower associated with the Giacobini-Zinner comet. The results on that occasion are shown in figure 5, together with the results for the control days, for the same night hours, about the same period. It will be seen from these curves that on the control days, there was the usual increase of ionization-bursts detected during the night; but on the occasion of the meteor shower there was an enormous increase during the period 0300 to 0400 G.M.T. For the curve shown in figure 5 the burst-frequency rate was calculated by taking rather long time intervals of one hour. By taking the much smaller time interval of one minute, the detail of the variation of rate is disclosed. In figure 6 is therefore shown how the rate of bursts, estimated in this way, varied with time throughout the shower.

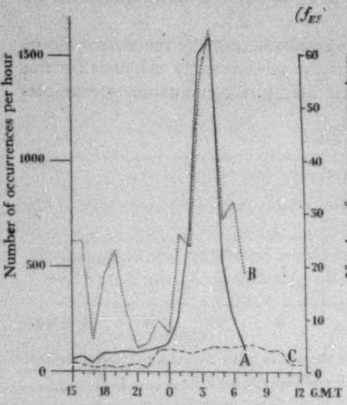


Figure 5. Showing variations of (A) ionization burst frequency, and (B) sporadic E-layer ionization during the night of the Giacobinid meteor shower. (A corresponding curve of ionization burst frequency (C) for a normal night is also shown.)

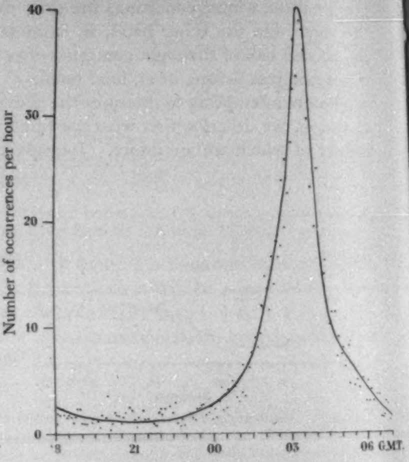


Figure 6. Showing detailed variation of ionization burst frequency during the Giacobinid meteor shower of 9/10 October 1947.

In plates A and B are shown some sample records for various times during the night of 9/10 October. These clearly show the remarkable display of ionization bursts which accompanied the meteor shower.

§6. DISCUSSION OF EXPERIMENTAL RESULTS: PART I (*continued*)

We now turn to discuss the results we have obtained, so far, on the subject of ionization bursts. There can be no doubt from the experiments of 10 October that meteor trails do give echoes of the type we have called ionization bursts.

It therefore remains to be considered whether all such bursts can be attributed to meteors. In this connection it is important to have regard to

(*a*) the diurnal and seasonal incidence of sporadic meteors;
(*b*) the exceptional influence of particular meteor showers; and
(*c*) the possible diurnal and seasonal variation of the response of the atmospheric medium.

In this connection it is well to recall that visual observations of meteors, are, of course, only possible during the hours of night-time; and the moon, when nearly full, makes such observations well nigh impossible. Such observations show, on the average, both a diurnal and seasonal variation in the number of visible meteors. The average hourly number after midnight is about double the average number before midnight. This is due to the fact that, in the evening, the meteors must overtake the earth, whereas in the morning hours we see both the meteors which the earth overtakes and those coming to meet it. Also meteors are twice as numerous from July to January as from January to July. These remarks refer to sporadic meteors which are not associated with any recognized shower.

In general it will be seen that our results fit in with what we should expect according to the meteor theory in that, during the night, the number of ionization bursts after midnight is approximately twice what it is before midnight. Concerning the annual variation it will be seen that the maximum in the curve of ionization bursts does not occur in the second half of the year, but in summer. The September values are, however, definitely higher than the March values. It appears unlikely that the curve in figure 1 indicates the true seasonal variation of sporadic meteors at noon, and we suggest that the variation disclosed indicates some form of solar control of the "response" of the atmospheric medium. For example, let us suppose that the "response" of the medium to the ionizing particles varies in some such manner as does the ionization in the E-layer, there being a maximum in summer and a minimum in winter. If, then, the incidence of the ionizing agency were greater in the second half of the year than in the first, the general shape of the resultant curve of detected ionization bursts might be expected to be as shown in figure 1.

§7. EXPERIMENTAL RESULTS: PART II

As mentioned above, we have relied for our results on abnormal or sporadic E-layer ionization on the Slough daily (h', f) ionospheric records. We have found that, generally, these records for the years 1945 and 1946 show results similar to those we have previously obtained and described (Appleton, Naismith and Ingram, 1939; Appleton and Naismith, 1940). Sporadic E-layer ionization is found to be greater in summer than in winter and greater by day than by night. But, by more accurate measurements, we have proved, more definitely than before, that abnormal E-layer ionization is controlled, at any rate in the daytime, in its magnitude and level of occurrence, by the normal E-layer. For example, in figure 7 is shown the diurnal variation of height of abnormal E-layer reflections, from which it will be seen that the level varies in exactly the same way as does the level of maximum E-layer ionization during the daytime. Moreover, we have found that the seasonal variation of average height at noon varies as we should expect if it is associated with the level of the normal E-layer. This is illustrated in figure 8.

It has been shown by Skellet and others that there is a temporary increase in abnormal E-layer ionization during periods of meteoric showers, and our own observations, illustrated in figure 5, where both ionization bursts and abnormal E-layer ionization are plotted for the night of 10 October, illustrate this same effect in a most striking way. But abnormal or sporadic E-layer ionization occurs at times other than those of meteor showers, and the real question at issue is whether or not we can consider meteors as one of the major sources of sporadic E-layer ionization, bearing in mind its very striking diurnal and seasonal characteristics which indicate some form of solar control.

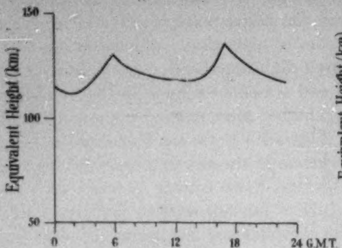


Figure 7. Showing diurnal variation of the average height of sporadic E-layer ionization.

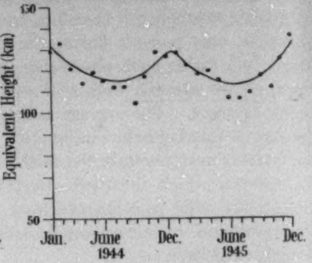


Figure 8. Showing seasonal variation of the average height of sporadic E-layer ionization at noon.

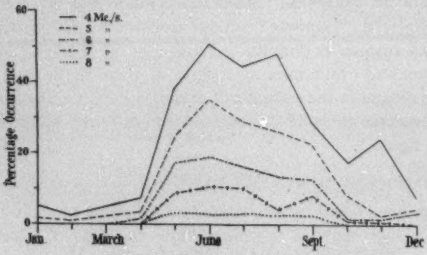


Figure 9. Percentage occurrence of sporadic E-layer reflections at noon at Slough on fixed radio frequencies (average for 1943–6).

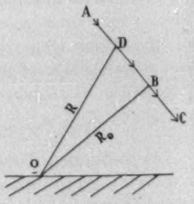


Figure 10.

The results described above, however, indicate precisely the same kind of solar control for ionization bursts, which we now can definitely attribute to meteors. The parallelism between the two phenomena of sporadic E-layer ionization and meteor-trail incidence is well illustrated in figure 11. The annual variation of sporadic E-layer ionization at noon is, perhaps, better illustrated in figure 9 where the percentage occurrence of sporadic E-layer reflections at noon at Slough for various frequencies is shown. The asymmetry of these curves relative to midsummer is clearly marked.

We therefore suggest that there are strong grounds for accepting sporadic meteors as being responsible in substantial measure for sporadic E-layer ionization in temperate latitudes. It should, however, be noted that, taking the earth as a whole, it has already become clear that there must be at least two sources of sporadic E-layer ionization. In the course of the International Polar Year investigations (Appleton, Naismith and Ingram, 1939) it was found that there was, in high latitudes, a high positive correlation between sporadic E-layer ionization and magnetic activity. We can therefore conclude that near the auroral belt, sporadic E-layer ionization can, to a large extent, be attributed to the cosmic agency (e.g. charged solar particles) responsible for the causation of magnetic storms.

On the other hand, at the lower latitude of Slough ($51\frac{1}{2}$° N.) there is no marked connection between sporadic E-layer ionization and magnetic activity, and the diurnal variation of such ionization has entirely different characteristics from those exhibited in high latitudes. A different originating agency is therefore suspected.

§ 8. SUMMARY OF CONCLUSIONS

It may be convenient to summarize briefly the results and conclusions emerging from the above discussion.

(*a*) Earlier conclusions, that transient radio echoes from the level of the E-layer are due to reflections from the ionization trails of meteors, are confirmed. The effective scattering area of such trails falls off as the sounding radio-frequency is increased. Thus, if high radio frequencies are used, only the denser trails (e.g. in meteor showers) are detected. With lower radio frequencies, the general incidence of sporadic meteors may be studied.

(*b*) There is a marked seasonal variation of the number of sporadic meteors recorded per hour at noon, the number being a maximum in summer and a minimum in winter. The curve illustrating this relation is not, however, symmetrical about midsummer, for it is found that more meteors are recorded in the second half of the year than in the first. It is not considered that this curve necessarily gives a correct indication of the seasonal variation of the actual incidence of meteors at noon, and a possible seasonal variation of the "response" of the atmospheric medium to the ionizing meteroric particles is suggested.

(*c*) Average curves of the diurnal variation of recorded ionized meteor trails fit in with the known features of the incidence of sporadic meteors so far as night-time conditions are concerned. The daytime portion of the diurnal curve of meteors recorded by radio methods is, however, found to alter with the season, most probably because the daytime "response" of the medium to the ionizing meteoric particles (referred to in (*b*) above) is greater in summer than in winter.

(*d*) During the Giacobinid meteor shower of 10 October 1946, a marked increase of sporadic E-layer ionization was found to accompany the very marked increase of ionized meteor trails recorded. This, and the striking parallelism between the daytime incidence of sporadic E-layer ionization and meteor ionization bursts, have led the authors to advance the theory that the fine dust of ionizing meteors is mainly responsible for the production of sporadic E-layer ionization in temperate latitudes.

(*e*) Measurements of the heights of occurrence of meteor ionization bursts and sporadic E-layer ionization indicate that, at any rate in the daytime, the former occur at a slightly lower level than does the latter. The atmosphere is, therefore, pictured as being bombarded with meteors and especially with dust particles too small to produce visible meteor trails. As these particles strike through the E-layer they cause additional ionization which, in the daytime, is detected by the usual methods of ionospheric sounding at the level of maximum E-layer ionization. During the night, sporadic E-layer ionization is detectable over a greater range of heights.

APPENDIX

A note on whistling meteors

Some years ago Chamanlal and Venkataraman (*Electrotechnics*, No. 14, Nov. 1941) described some very interesting experiments in which they observed a radio Doppler effect due to meteors entering the earth's atmosphere, which was exhibited in the form of whistles heard in the telephone of a radio set tuned to an unmodulated short-wave radio station. The observed whistles were found to descend in pitch, the audible note usually continuing to zero frequency, though on occasion the note ceased before zero frequency was reached. The explanation of the phenomenon given by Chamanlal and Venkataraman was that the whistles were heterodyne notes caused by the interference of the direct radio wave from the sending station and the reflected wave from the rapidly moving ionized surface associated with the meteor. The same authors further stated that "the descending pitch of the heterodyne beat note is the result of the moving reflecting surface being rapidly retarded in velocity, since it is evident that the beat note will reduce to zero frequency if the velocity of the reflecting surface becomes zero."

In our own observations on meteor reflections we have often observed whistles of falling frequency of this kind, and, through the kindness of Mr. L. W. Hayes of the B.B.C., we have been furnished with observations on the same subject made at the B.B.C. Listening Station at Tatsfield.

As a result of our attempt to correlate the results obtained by the two methods of meteor detection, by pulses and by Doppler whistles, we have been led to question the validity of the explanation of the varying pitch of the whistles, due to Chamanlal and Venkataraman, and quoted above. We may illustrate the matter with reference to figure 10 where the simplest type of experimental arrangement is depicted and where O may be taken as the site of both sending and receiving stations (pulse and C.W.).

Let us suppose that the meteor trail is ADBC and that the meteor velocity is v. Further, let t be the time when the meteor is at D, and thus at range R; and let t_0 be the time when the meteor is at B, the minimum distance. Let this minimum distance be R_0. During the period when the meteor is travelling from A to B we have

$$R^2 = R_0^2 + v^2(t_0 - t)^2, \qquad \text{......(1)}$$

and thus

$$\frac{dR}{dt} = -v\left(1 - \frac{R_0^2}{R^2}\right). \qquad \text{......(2)}$$

For the observations on continuous waves, the beat frequency observed at 0, due to the interference between the direct wave and the wave reflected from the meteor trail, is given by

$$f=\frac{2}{\lambda}\frac{dR}{dt}=-\frac{2v}{\lambda}\left(1-\frac{R_0^2}{R^2}\right)^{\frac{1}{2}}. \qquad \ldots\ldots(3)$$

Thus as the meteor particle travels from A to B the value of f should fall and this is found to be the case. The frequency will become zero when R is equal to R_0. We thus see that the fall of the whistle frequency can be accounted for without assuming retardation of the meteor itself. The insertion of typical values in the right-hand equation of (3) gives values of f of the order actually measured.

Now Pierce (1938) has emphasized the fact that the strongest reflection may be expected from a meteor trail which is broadside on to the exploring radio station. In other words, the strongest reflection received back at 0 will come from B on the meteor trail. Experimental evidence supporting this has been given by Hey and Stewart (1946). In our own experiments using radio pulses, we believe that it is usually the distance R_0 which is measured, since the value of the range remains fairly constant during the period when the echo is received.

(It may be of more than theoretical interest to note that, if a record of whistle frequency with time could be obtained, together with a pulse determination of the range R_0 for the same meteor, it would be possible to find the meteor velocity v. For, from (3), we have

$$R=R_0+\frac{\lambda}{2}\int_{t_0}^{t} f\,.\,dt, \qquad \ldots\ldots(4)$$

so that the value of R could be inserted in the right-hand side of (3) and the value of v found.)

It will be seen that, according to the explanation of the connection between meteor whistles and meteor pulses echoes given above, the former phenomenon occurs earlier than the latter. Now, it has been frequently reported by observers of meteor whistles, listening within the "skip" distance to a near-by short-wave sending station, that the whistles are succeeded by a burst of signal strength. We think that the burst of signal strength corresponds to the arrival of the meteor head at the broadside-on position B (see figure 10) and so corresponds in time with the reception of the strong pulse echo.

ACKNOWLEDGMENT

The work described above was carried out as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research.

REFERENCES

Appleton, 1931. *Nature, Lond.*, **127**, 197.
Appleton and Naismith, 1932. *Proc. Roy. Soc.*, A, **137**, 36.
Appleton and Naismith, 1935. *Proc. Roy. Soc.*, A, **150**, 685.
Appleton and Naismith, 1940. *Proc. Phys. Soc.*, **52**, 402.
Appleton and Piddington, 1938. *Proc. Roy. Soc.*, A, **164**, 467.
Appleton, Naismith and Ingram, 1937. *Phil. Trans. Roy. Soc.*, A, **236**, 191.
Appleton, Naismith and Ingram, 1939. *Proc. Phys. Soc.*, **51**, 81.

BHAR, 1937. *Nature, Lond.*, **139,** 470.
HEY and STEWART, 1946. *Nature, Lond.*, **158,** 481.
MITRA, SYAM and GHOSH, 1934. *Nature, Lond.*, **133,** 533.
MOHANTY, 1938. Ph.D. Thesis (Cambridge).
PIERCE, 1938. *Proc. Inst. Rad. Engrs.*, **26,** 892.
SKELLET, 1931. *Phys. Rev.*, **37,** 1668.
SKELLET, 1935. *Proc. Inst. Rad. Engrs.*, **23,** 132.

DISCUSSION

Mr. G. R. M. GARRATT. By the courtesy of Mr. Cecil Goyder, Chief Engineer of All India Radio, I was privileged a few months ago to observe the phenomenon of " whistling meteors " in New Delhi and to meet the two members of his staff, Messrs. Chamanlal and Venkataraman, who share the credit for this very interesting discovery.

I believe it was during the winter of 1940/41 that this effect was first observed. Whistles were heard while monitoring the short-wave transmitters of All India Radio which could not be explained as ordinary heterodynes. The whistles were of short duration, normally from about 2 to 3 seconds; each commenced as a note of about 3000 cycles, fell rapidly in pitch to zero or sometimes died away before reaching zero, and only in rare cases did the whistle reappear as an ascending note on the other side of zero.

It was eventually deduced that these unusual characteristics could only be explained on the assumption that they were due to a Doppler effect resulting from the interference between the weak ground waves received direct from the transmitter about 9 miles away and waves reflected from some very rapidly moving surface. The velocity of the reflecting surface towards the observer was shown to be of the order of 40–80 km./sec. and it seemed that the only likely source of such high velocities would be a meteor producing ionization in the earth's upper atmosphere. This assumption was easily confirmed in the clear sky of Northern India where " shooting stars " can be seen on almost any night of the year by establishing direct correlation between the occurrence of a whistle and the arrival of a visible meteor.

Although the whistles can be detected by any sensitive receiver capable of receiving broadcasts on the short wave bands, there are a number of conditions which must be satisfied before success will be achieved. Firstly, it is necessary to have a high power transmitter—preferably not less than 50 kw.—radiating an unmodulated signal on a frequency of the order of 5–15 Mc./s. The receiver must be situated within the skip distance so that no reflections are received from the E- and F-layers, yet it must be so located that the direct ground wave is a weak one. If the ground wave received is too strong, the normal operation of the A.V.C. circuits cuts down the amplification of the receiver and prevents detection of the very weak reflected waves. There is thus an optimum strength of ground wave which, with a normal receiver, is of the order of 70 to 100 mv. It has been estimated that the reflected signal from the meteor may be as small as 0·5 mv. although in the case of a meteor passing near the zenith, the energy received may be several hundred times as great as this.

Even in the clear sky of Northern India, only a small number of meteors are visible to the eye compared with the number which can be heard. The ratio is probably of the order of 100 : 1—though any meteor which is visible can be relied on to produce a prominent whistle.

These effects can be heard on almost any night of the year—the greatest number being heard about 4.0 a.m. The number to be heard on an ordinary night, however, may be only of the order of 4–6 per minute as compared with an almost continuous performance at times of meteoric showers.

It has been stated that the " whistle " is invariably characterized by a falling note, which has been attributed to the loss of velocity of the meteor as it is retarded in the earth's atmosphere. Although it is true that the great majority of whistles exhibit a falling note (which often dies away before reaching zero frequency), a small number yield a note which falls in pitch, passes through zero, rises again and then dies away at about 1 to 2000 cycles. A very small number seem to show only a rising note.

I think it is clear that these effects are due to a combination of circumstances. A Doppler effect or frequency shift arises from the component of velocity towards the observer and if the reflecting surface had a fixed component of velocity towards the observer, a beat note of

PROC. PHYS. SOC. VOL. 59, PT. 3 (SIR E. APPLETON AND R. NAISMITH)

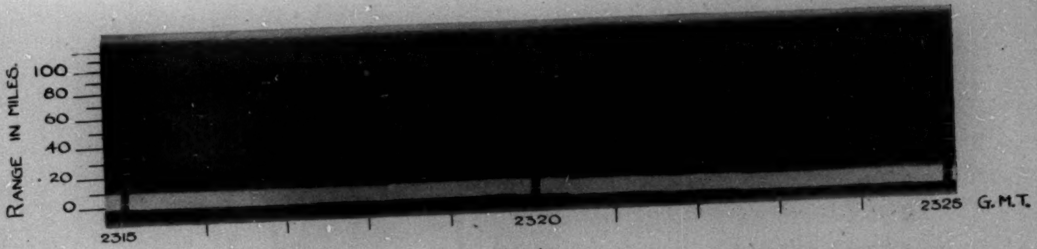


Plate A.

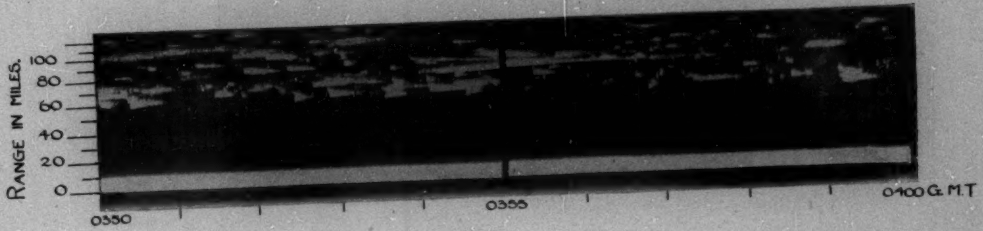


Plate B.

Ionization trails recorded at Radio Research Station, Slough. Plate A on 9.10.46; plate B at the height of the Giacobinid meteor shower on 10.10.46.

To face page 472

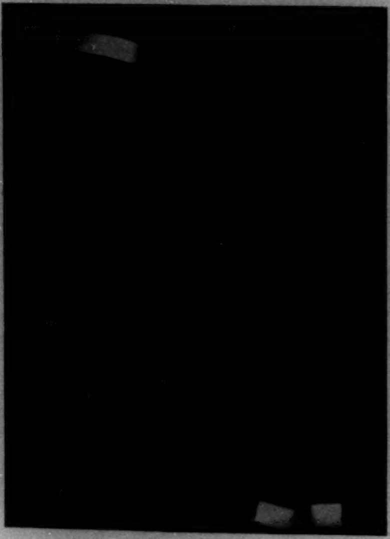


Figure 1. Illustration of the type of trace obtained (positive).

constant frequency would be heard. A falling note can be produced either by a retardation of the meteor in the upper atmosphere, in which case the component of velocity towards the observer also diminishes, or it can be produced by a meteor travelling at constant velocity in such a direction that its track lies tangentially with respect to the observer within the somewhat restricted region in which reflected echoes can be detected. For example, a meteor travelling horizontally and passing directly through the observer's zenith has a component of velocity towards the observer which diminishes to zero as it passes overhead and then increases away from him. Such a meteor travelling at constant velocity will give rise to a descending note as it approaches the zenith and an ascending note as it recedes.

In practice, the beat note heard must be almost invariably a combination of both effects, the retardation of the meteor in the atmosphere and the changing component of velocity with respect to the observer. Since meteors are retarded very rapidly in the earth's atmosphere, however, it is generally only those travelling on an approximately horizontal course at extreme altitudes where there is little retardation which have any appreciable velocity left after passing the zenith to produce a rising note at the receiver.

The mechanism by which a meteor gives rise to a beat note of rising frequency only, without a preceding fall through zero frequency, is not very clear. It seems possible, however, that it is not unconnected with the fact that the transmitter and receiver are situated some miles apart and the response received may therefore depend to some extent on the direction of the reflecting surface in relation to that of the transmitter.

Finally, it may be useful to note that if the meteor has a component of velocity towards the observer, the heterodyne whistle appears on the higher-frequency side of the ground wave. If the meteor is moving away, however, the whistle appears on the lower side of the ground transmission, and if it is specially desired to observe the somewhat rare cases where a rising note is exhibited, it is advisable to detune the receiver slightly on the low-frequency side, since the whistle in these cases is always a weak one and may otherwise be missed altogether.

THE SHORT-PERIOD TIME VARIATION OF THE LUMINESCENCE OF A ZINC SULPHIDE PHOSPHOR UNDER ULTRA-VIOLET EXCITATION

By MARY P. LORD,* A. L. G. REES† and M. E. WISE,
Philips' Lamps Ltd., Mitcham, Surrey
* Now at Imperial College, London
† Now in Melbourne, Australia

MS. received 20 *February* 1946 ; *in revised form* 18 *October* 1946

ABSTRACT. A photographic method, in some respects new, is described for measuring in detail the luminescent-intensity time relationships of a powdered phosphor. A zinc sulphide with silver and copper impurities was irradiated at 20, 40, 60 and 80° c. with various intensities of λ3650 : the results are mainly on the blue phosphorescent intensity over some 300 milliseconds from the end of irradiation.

The observations are found to correspond quantitatively to a bimolecular law with two types of activating centre. On this assumption four independent constants are deduced mathematically for each curve, two per centre. They are interpreted in terms of the initial concentrations of ionized centres, and the recombination coefficient for each kind of centre with free electrons.

Using these constants, the ratio of the contributions from the two kinds of centre to the luminescent intensity at the beginning of the decay was deduced. From this ratio and the spectral distribution of the fluorescent emission, the centre with the faster decay was

identified as silver and the other as interstitial zinc. The very slowly decaying green band present in the luminescent emission is ascribed to the presence of copper centres.

The constants are found to change with temperature and with exciting intensity in a way which cannot be explained on the simple theory. These variations can, however, be accounted for by assuming that there are trapping centres closely associated with each "excited" activating centre. It is shown that the traps associated with the copper centres are of major importance, although the concentration of copper impurity is relatively small.

As functions of exciting intensity the constants also have periodic terms superimposed on the main changes. An attempt is made to relate this qualitatively to the equations for the build-up of luminescence when more than one type of activating centre is present.

The effect of the thickness of the layer of powder on the simplest type of two-centre bimolecular decay law and the relation, in outline, of our interpretation to the work of other observers are also considered.

§1. INTRODUCTION

THE mechanism of the luminescence of inorganic solids has been the subject of much theoretical and practical investigation for many years, but in spite of the stimulus to this research provided by recent technological applications, the problem has remained very far from solution.

An experimental method has been evolved which permits observations to be made at time intervals of the order of 4 milliseconds over a range of 400 milliseconds from the beginning of the decay. Thus the results may be more detailed than any previously published over the range investigated.

§2. EXPERIMENTAL

The principle of the method used has been applied in many ways by earlier workers. A well-defined beam of monochromatic ultra-violet radiation strikes normally a layer of the luminescent material spread evenly on a glass disc that can be rotated uniformly at an appropriate speed about its axis. The light from the phosphor is recorded photographically; and, from the density of blackening at various points of the trace and the plate characteristics, the intensity of the phosphorescent light may be obtained.

The whole apparatus was mounted on a substantial base. Exciting radiation was obtained from the quartz inner tube of an MBF/V. 80-watt mercury-discharge lamp operated on d.c., the power consumed being controlled by a rheostat and measured on a Hartmann and Braun wattmeter. The radiation from the lamp was passed through a Wratten 18 A filter to ensure that approximately monochromatic radiation of wave-length 3650 A. was incident on the phosphor. The beam was collimated as far as possible by passing through a series of limiting apertures, the final one being just clear of the phosphor and shaped as a segment (20°) of an annulus having bevelled edges. The excitation time in minutes was, therefore, 20/360 × speed of the disc (coated with phosphor) in revolutions per minute. The disc is driven through a gear train of 2 : 1 ratio by a specially adapted generator (12 volts d.c. input, 100 volts 1 amp. output). The speed of the generator could be varied by means of a resistance (1–10 ohms) in the driving commutator circuit; the loading of the secondary was provided by a low-resistance voltmeter whose reading was proportional to the speed of the motor when the field current was maintained constant. The absolute speed was determined both by a stroboscopic method and by direct counting, and it was found that uniform speeds of rotation were obtained. The glass disc was coated by spraying on to a thin adhesive layer; the combined thickness of the adhesive and phospher layers (measured by an interference method) is *c.* 0·08 mm. and the particle size distribution of the phosphor is

Diam. (μ)	8	8–16	16–23	23–32	32–45	45–64
No. (%)	60·4	20·6	10·6	5·2	2·3	0·9

To enable observations to be made at temperatures up to 100° c., a thermally-insulated and light-tight jacket was fitted over the rotating disc. Into a circular aperture directly opposite the disc was fitted the lens of a quarter-plate field camera used at double extension. All joints were lined with light-proof black felt so that no extraneous light could enter the camera. Filters could be incorporated behind the lens in the camera; throughout this series of experiments a sodium nitrite solution was used to remove any λ3650 radiation which had entered the camera and a Wratten No. 39 filter to remove a faint green band occurring in the luminescent emission. Half of the disc could be photographed; each trace was produced by many (of the order of 100) revolutions of the disc, the first few activations being in no case recorded. Ilford H.P.3 plates were used throughout and a standard procedure adopted for their development. Using previously calibrated neutral filters, the plate characteristics were determined directly for the equilibrium fluorescence on each plate, thus eliminating the effect of slight differences in development conditions. The characteristics of H.P.3 plates at several wave-lengths in the range of the luminescent emission were also measured; using these, it was possible to allow for the effect of changes in the spectral distribution of the luminescent emission during build-up and decay. To avoid errors due to failure of the reciprocity law, the exposure time was made equal to that of the corresponding trace. By photographing a strip of standard intensity on each plate, the intensities calculated for the different plates were obtained in terms of a common scale factor. A representative positive of the type of trace obtained is illustrated in figure 1. To measure the transmission of the circular trace, a Hilger microphotometer was adapted for use with a rotating stage. Measurements could then be made at intervals of 30 minutes of arc, corresponding to a time interval of 0·00139 sec. for a speed of 60 revolutions per minute. The usual precautions were taken to ensure constant intensity from the microphotometer lamp; the relationship between light intensity and response of the photocell galvanometer combination was linear. Density determinations were made at a number of intervals around the trace and converted to luminescent intensity values.

The method outlined here has several advantages over those employed previously. Most previous investigators (see, for example, Beese (1939)) employed direct measurement on a rotating disc or cylinder coated with phosphor. The slit-widths used with the photocells covered a considerable time range, a factor which automatically introduces a large error, particularly at the shorter times where the intensity is changing rapidly. The electron multiplier-cathode ray oscillograph method employed by de Groot (1939a) and others is only suitable for accurate measurements over a period of 20 milliseconds or so.

Furthermore, the method outlined here has the advantage of furnishing a permanent record of the phosphorescence.

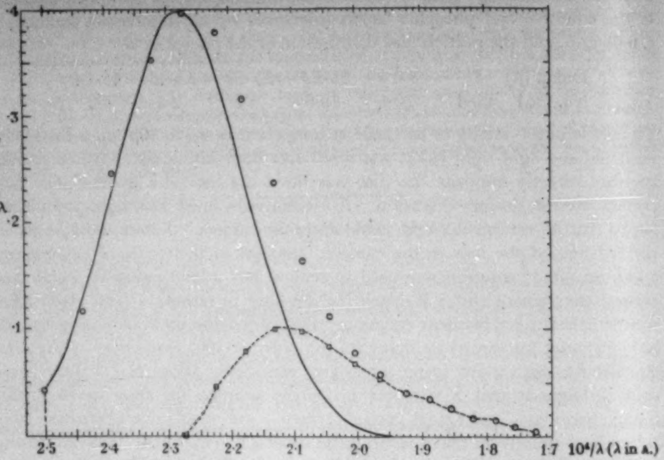


Figure 2. Spectral energy distribution of silver-activated zinc sulphide. ⊙ Experimental points.

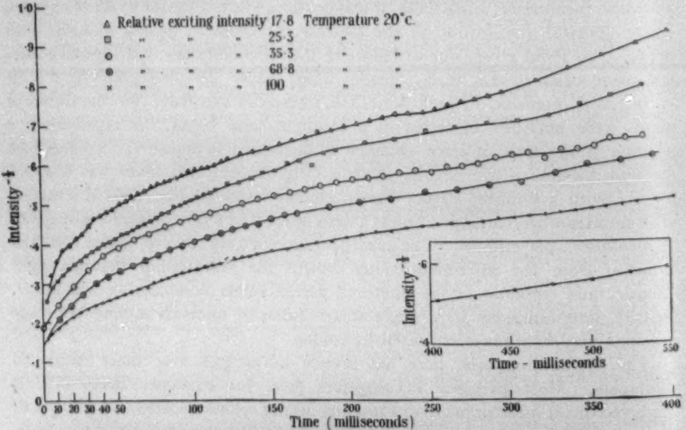


Figure 3. Plots of $y^{-\frac{1}{2}}vt$ for various exciting intensities and temperatures.
(*a*) Temperature 20° c.

§ 3. RESULTS

The crystal phosphor used in these investigations was a zinc sulphide, activated by *c.* $7 \cdot 5 \times 10^{-3}$ % silver, but containing also some ($\leqslant 1 \times 10^{-4}$ %) copper impurity, which caused further independent activation, whilst small quantities of iron ($\leqslant 15\times10^{-5}\,\%$), nickel ($\leqslant 2\times10^{-5}\,\%$) and chlorine ($\leqslant 150\times10^{-5}\,\%$) were probably also present. The spectral energy distribution of the light emitted by the phosphor when excited by almost monochromatic radiation of wave-length 3650 A. is given in figure 2. The significance of the overlapping continua present in this distribution will be discussed later.

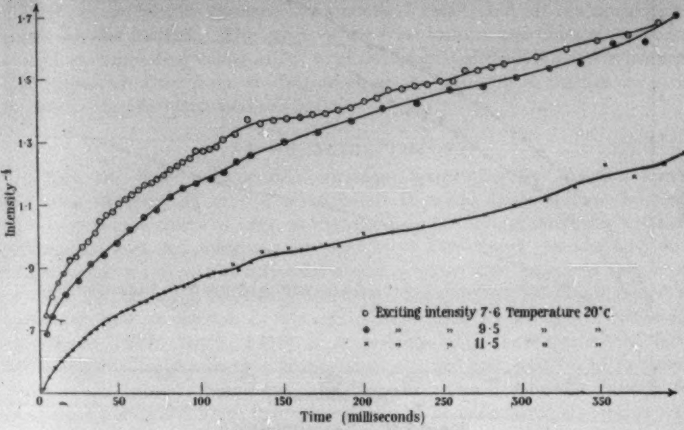


Figure 3 (*b*). Temperature 20° c.

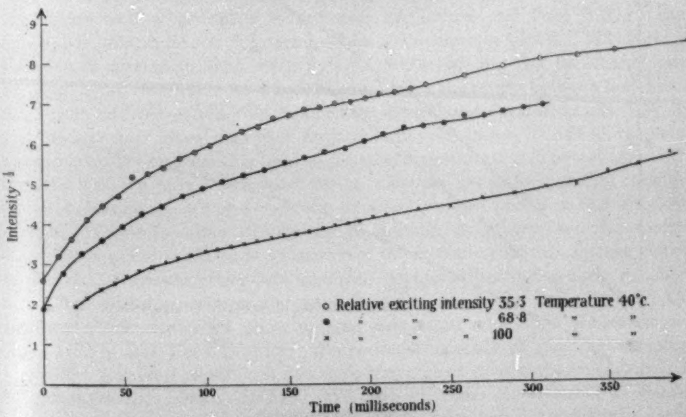


Figure 3 (*c*). Temperature 40° c.

The blue luminescent intensity-time relationships were measured over a range of exciting intensities at temperatures from 20° c. to 80° c. At some of the higher temperatures there were periods in which the density of blackening

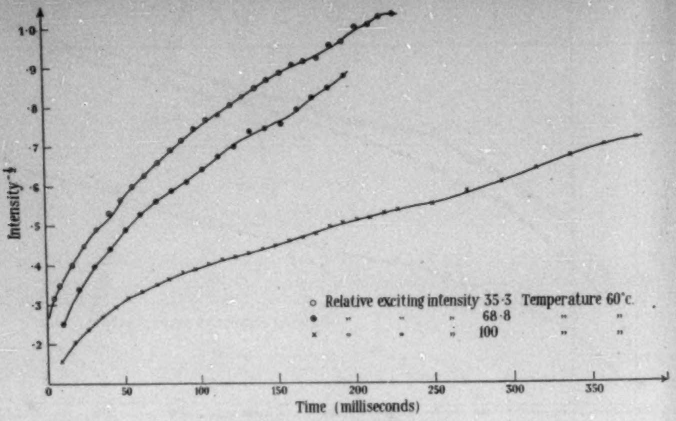


Figure 3 (*d*). Temperature 60° c.

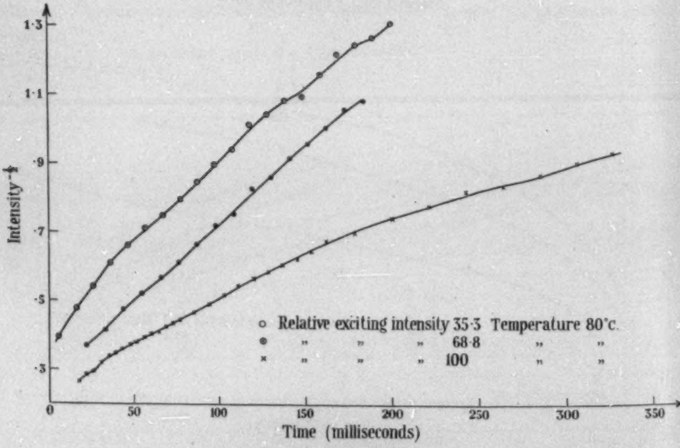


Figure 3 (*e*). Temperature 80° c.

of the photographic plate was too great to measure, so that the corresponding observations are missing. Owing to the smallness of the density of blackening at the longer times, the corresponding intensities may be unreliable. To diminish the random errors, all the observations except some extreme ones were smoothed by the least squares formula, which uses seven consecutive results to smooth the middle one of the seven. In order to test the theory of § 5, the $-\frac{1}{2}$ powers of the phosphorescent intensity are plotted against time in figure 3. The curves are drawn through the smoothed values and all the unsmoothed points are also plotted. The range of the observations, the times of activation and the corresponding values at 20° C. of the integral of the luminescent intensity with respect to time from the beginning of the build-up to infinity are given in table 1; they are discussed in § 5.

§ 4. PHYSICAL THEORY

There has been considerable argument concerning the function of the activating impurity in crystal phosphors. It seems fairly certain, however, that the activator atoms or ions, as the case may be, occupy interstitial positions in the crystal and that absorption of radiation of wave-length greater than 3340 A. in zinc sulphides occurs at intestitial atoms, whether they be excess zinc atoms or those of the added activating impurity (Seitz, 1938; 1939 **a** and b). The suggestion that an exciton is formed by absorption by a negative lattice ion (Kitchener, 1939; Riehl, 1939) is at variance with the experimental facts. Milner (1939) has shown that absorption by lattice ions would lead to a band width much greater than 1 ev. Furthermore, the quantum efficiency for irradiation with λ 3650 is *c.* 1, a value hardly to be expected if absorption takes place at lattice ions some distance from the source of fluorescent emission, namely the interstitial atoms. Moreover, evidence that electrons are excited to the conduction band is provided by the work of Gudden and Pohl (1921): they found that photoconductivity accompanies absorption by λ 3650. We therefore consider, in agreement with Seitz (1938, 1939 **a** and b), that the energy level occupied by an electron in an activator atom lies in the region between the highest filled band and the conduction band of the crystal at a point determined by the type of activator atom, and that excitation by absorption of λ 3650 causes an electronic transition from this level to the conduction band. There is agreement that the positive hole left behind by an electron generally remains localized at the activating centre and that the emission process occurs at the activating centres. There is some divergence of opinion as to whether the spectral distribution of the emission is a function of the activating centre, or of the matrix, or of both. However, silver-activated zinc sulphides often have a band maximum at 4400 A. and copper-activated ones, a green emission. Both Ag- and Cu-activated ZnS sometimes show a band maximum at 4650 A. (Leverenz and Seitz (1939) and Rees (1942)); the emission spectra of pure zinc-sulphide phosphors also have maxima at this wave-length, and their emission is attributed to interstitial zinc (Seitz, 1939 **a**). We have therefore endeavoured to identify the centres in "our" phosphor by analysing the relationship between the energy ($E_\lambda d\lambda$) emitted in the wave-length range $\lambda \rightarrow \lambda + d\lambda$ and the wave-number ($1/\lambda$). This relationship (see figure 2) has a maximum at *c.* 4400 A. and a long wave-length tail extending to *c.* 5800 A. Over the range 4000–4400 A., $\lambda^2 E_\lambda$ is Gaussian, and hence, following Henderson (1939), we assume that this part is due to a single centre which we identify as silver. It can be seen from figure 2 that subtraction

Table 1. Times of activation t_0, values of $\int X\,dt$, and estimated values of $\dfrac{\int^{\infty} X\,dt}{I t_0}$ over build-up and decay

(I denotes exciting intensity; t denotes time of last observation)

		$\int X\,dt$. Temperature 20° c.							Temperatures (°c.)					
									40°		60°		80°	
I	t_0	Over build-up	Over observed decay	Remainder (extrapolated from $1/b_1(b_1t+C_1)$)	Total	$\dfrac{\int^{\infty} X\,dt}{I t_0}$	t	I	t_0	t	t_0	t	t_0	t
100	96·0	3498·5	4000	3136	10635	1·108	548	100	79·4	449	87·6	381	85·4	329
68	70·8	1956	2535	2843	7334	1·524	452	68	122·4	308	102·0	194	108·8	184
35·3	70·7	1164·5	1750	1829	4744	1·902	384	35·3	94·4	401	80·4	225	103·0	201
25·3	79·0	750·8	1425	1353	3529	1·579	384							
17·8	78·0	602	1062	1123	2787	2·008	464							
11·0	78·0	200·3	488	643	1331	1·552	403							
9·54	74·0	111·1	270	365	746	1·057	462							
7·20	79·0	91·2	234	364	689	1·211	388							

N.B.—Unit of time is 1 millisecond.

of the silver band * from the complete distribution leaves an asymmetrical band in which *c.* 20% of the total energy is emitted. This complex consists mainly of a single band with a maximum at *c.* 4680 A., and there is also a broad band with a maximum in the green. We therefore consider the phosphor to contain three types of centre: Ag, Zn and Cu in order of increasing wave-length of their maxima.

There is a divergence of opinion about the history of the electron between the time when it is excited into the conduction band and the time when it returns to an activating centre. Many workers (e.g. Lewschin and Antonow-Romanowsky (1934)) consider that the time is spent in wandering freely through the conduction band, and have shown that such behaviour leads to a bimolecular type decay. Randall and Wilkins (1945 b) and Garlick and Wilkins (1945) have, however, pointed out that for decays of the duration commonly found, this theory would lead to rather small values of the cross-section of capture of an electron by a luminescence centre; also, that the decay rate would be independent of the time of activation, and would increase slowly with temperature, and that there would be an inflexion in the build-up curve. They obtained curves of this type for some phosphors over a limited temperature range, but most of their observations do not support the simple biomolecular theory. They suggest that the electron spends the greater part of its lifetime in the excited state in a trap (or traps) from which it can be released by absorbing thermal energy from the lattice: the number and depth of the traps are considered to be a function of the structure of the phospher only.

However, we have found that, assuming two types of activating centre, a bimolecular law can be fitted to each of our decay curves with the exception, in some cases, of a few points at the beginning and end. The variation with temperature and exciting intensity of the four independent constants obtained from each curve cannot be explained on either the simple bimolecular theory or on a theory incorporating the ideas of Garlick and Wilkins (1945) on bimolecular type decays. Nearly all our results are, however, explained by introducing a new trapping mechanism.

§ 5. NUMERICAL ANALYSIS OF THE DECAY RESULTS ASSUMING THAT THERE ARE TWO TYPES OF ACTIVATING CENTRE (SILVER AND ZINC)

We shall denote by n_1 and n_2 the number of excited activating centres per unit volume of the first and second kinds.

The excited electrons wander freely through the conduction band. Luminescence is emitted when such an electron is captured by an ionized activating centre and falls to a lower energy level. The probability that a particular activating centre will be filled in unit time is proportional to the concentration of the electrons (of number n) per unit volume in the conduction

---

* Henderson (1939) has found that one kind of centre gives rise to a spectral distribution whose E_λ values (with the exception of the smallest ones) are well represented by the Gaussian expression

$$E_\lambda = \frac{A'}{\lambda^2\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}\left(\frac{10^4}{\lambda}-\frac{10^4}{\lambda_0}\right)^2}.$$

The experimental observations between 4000–4400 A. lead to $\sigma=0{\cdot}1046$, $\lambda_0=4386$ A. ($A'=1$) for the silver band of our phosphor.

band round the centre, and is assumed to be $\beta_1 n$ for an activating centre of the first kind and $\beta_2 n$ for an activating centre of the second kind.

Then for a unit volume of the lattice, within which we can neglect the effect of variations of n,* the equations for the decay are

$$dn_1/dt = -\beta_1 n_1 n, \qquad \ldots\ldots(5.1)$$

$$dn_2/dt = -\beta_2 n_2 n, \qquad \ldots\ldots(5.2)$$

$$n = n_1 + n_2. \qquad \ldots\ldots(5.3)$$

To apply these equations to the analysis of our observations, it is necessary to relate the observed luminescent intensity to the rate ($-dn_1/dt - dn_2/dt = X_1 + X_2 = X$, say) at which quanta are emitted by the two kinds of centre. If $(E_{1t\lambda}/h\nu)d\lambda$, $(E_{2t\lambda}/h\nu)d\lambda$ quanta per second per unit volume are emitted by centres of the first and second kind respectively at time t in the wave-length range $\lambda \rightarrow \lambda + d\lambda$, the recorded luminescent intensity (y) is $A'\int F_\lambda(E_{1t\lambda} + E_{2t\lambda})d\lambda$. A' is the same for all observations, and is the ratio of the recorded intensity of the standard strip to the value of the integral for this strip. F_λ is a ratio † depending on the transmission factor of the blue filter and the deviation of the characteristic corresponding to the spectral distribution for the standard strip from the characteristic at wave-length λ. The band spectra of the centres are unlikely to change shape during the time variation of the luminescence, so we can put

$$E_{1t\lambda} = \frac{X_1 E_{1\lambda}}{\int \frac{E_{1\lambda}}{h\nu} d\lambda}, \qquad E_{2t\lambda} = \frac{X_2 E_{2\lambda}}{\int \frac{E_{2\lambda}}{h\nu} d\lambda},$$

where $E_{1\lambda}$ and $E_{2\lambda}$ are independent of time and proportional to the heights at wave-length λ of the repsective blue bands to which they correspond. Then

$$y = A'\left[X_1 \frac{\int F_\lambda E_{1\lambda} d\lambda}{\int \frac{E_{1\lambda}}{h\nu} d\lambda} + X_2 \frac{\int F_\lambda E_{2\lambda} d\lambda}{\int \frac{E_{2\lambda}}{h\nu} d\lambda}\right]. \qquad \ldots\ldots(5.4)$$

This can be written as

$$Ay = -dn_1/dt - A_F dn_2/dt. \qquad \ldots\ldots(5.4')$$

A_F is found to be 2 if centres of the second kind are silver; it will, of course, be $\frac{1}{2}$ if centres of the second kind are zinc.

The equations (5.1)→(5.4′) can be solved completely. In discussing them it is convenient to introduce

$$n_{10} = \text{value of } n_1 \text{ at } t=0;\ n_{20} = \text{value of } n_2 \text{ at } t=0;\ \zeta = n_{20}/n_{10};$$
$$\rho = \beta_2/\beta_1;\ u = n_1/n_{10}.$$

---

* The effect of variations of n within the layer is considered in Appendix B.

† It is the product of the transmission factor of the filter and a response factor obtained from H.P.3 characteristics measured at various wave-lengths (§ 2); the response factor appears to be independent of intensity over a wide range of intensities. The relative values of F_λ are 38, 38, 43, 35, 9·2, 1·53 and 0 at λ=4000, 4200, 4400, 4600, 4800, 5000 and 5200 A. respectively.

We can immediately obtain n_2 in terms of n_1 by dividing (5.2) by (5.1):

$$n_2 = n_{20} u^{\rho}. \qquad \ldots\ldots(5.5)$$

Hence X and t can be obtained in terms of n_1 only:

$$X = \beta_1 n_{10}^2 u^2 (1 + \zeta u^{\rho-1})(1 + \rho\zeta u^{\rho-1}) \qquad \ldots\ldots(5.6)$$

and

$$Ay = \beta_1 n_{10}^2 u^2 (1 + \zeta u^{\rho-1})(1 + A_F \rho\zeta u^{\rho-1}), \qquad \ldots\ldots(5.6')$$

$$t = \frac{1}{\beta_1 n_{10}} \int_u^1 \frac{du}{u^2 + \zeta u^{(1+\rho)}}. \qquad \ldots\ldots(5.7)$$

We shall designate the centres so that $\rho > 1$. Then n_2/n_1 ultimately becomes negligible. If $\rho > 2$ it is best to integrate (5.7) by parts and write it as

$$\beta_1 n_{10} t = \frac{1}{u(1 + \zeta u^{\rho-1})} - \frac{1}{1+\zeta} - \zeta(\rho - 1)\left[\int_0^1 - \int_0^u \frac{u^{\rho-3} du}{(1 + \zeta u^{\rho-1})^2}\right]$$

$$= \frac{1}{u(1 + \zeta u^{\rho-1})} - C + \zeta(\rho - 1) S_u, \qquad \ldots\ldots(5.8)$$

where

$$S_u = \int_0^u \frac{u^{\rho-3} du}{(1 + \zeta u^{\rho-1})^2} \qquad \ldots\ldots(5.9)$$

and

$$C = \frac{1}{1+\zeta} + \zeta(\rho - 1) S_1. \qquad \ldots\ldots(5.10)$$

Thus S_u is a power series with terms $u^{\rho-2}$, $u^{2\rho-3}$, etc. So after a long time

$$\beta_1 n_{10} t + C = 1/u \qquad \ldots\ldots(5.11)$$

and

$$y^{-\frac{1}{2}} = b_1 t + C_1, \qquad \ldots\ldots(5.12)$$

where

$$b_1 = \sqrt{(\beta_1 A)} \qquad \ldots\ldots(5.13)$$

$$C_1 = \frac{C b_1}{\beta_1 n_{10}}. \qquad \ldots\ldots(5.14)$$

The linear relationship (5.12) is satisfied by the observations over a wide range of values of t (given in table 2 *a*) in all decay curves. Sometimes the last few values are too large at 20° C. and too small at 60° C.; these in any case are experimentally unreliable. b_1 and C_1 were thus obtained numerically (table 2 *a*).

The next stage was to test (5.6) and (5.7) at shorter times by including the $u^{\rho-1}$ terms. We can expand (5.8) in powers of $\zeta u^{\rho-1}$, and when this is small, but not negligible, there will be observations for which

$$y(b_1 t + C_1)^2 - 1 \doteqdot B(b_1 t + C_1)^{1-\rho}, \qquad \ldots\ldots(5.15)$$

where

$$B = \zeta\left(1 + \rho A_F + \frac{2}{\rho - 2}\right)\left(\frac{C_1}{C}\right)^{\rho-1} \qquad \ldots\ldots(5.16)$$

Log$[y(b_1 t + C_1)^2 - 1]$ was therefore plotted against $\log(b_1 t + C_1)$ for all observations (except the 80° C. ones, for which $y^{-\frac{1}{2}}$ did not deviate sufficiently from $b_1 t + C_1$). A specimen curve is given in figure 4. It was then clear that at 20°, 40° and 60° C. (5.15) was correct from *c.* 20 milliseconds onwards. ρ (table 2 *a*) was obtained from the slope of the lines; its values are probably not so accurate as those of b_1 and C_1 since the values of $y(b_1 t + C_1)^2 - 1$ are small differences. However, within these limits ρ has the same values for different exciting intensities, viz. 20 at 20° c., 17·5 at 40° c., and 11·5 at 60° c. This is new evidence for the two-centre theory.

The intercepts of the figure 4 type curves give $\log B$ numerically (table 2 *a*).

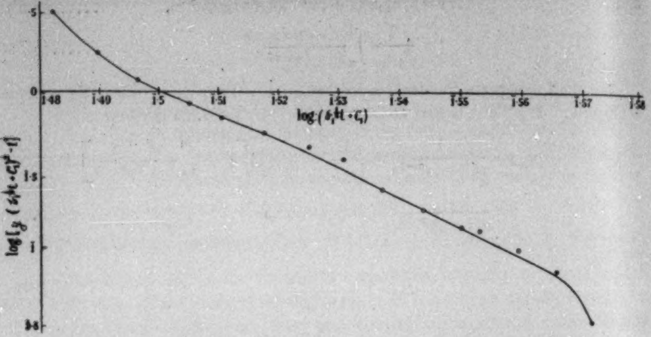


Figure 4. A representative plot of $\log\{y(b_1t+C_1)^2-1\}$ against $\log(b_1t+C_1)$.

Table 2 *a*. Constants obtained directly from the decay observations

Temperature (°c.)	I	b_1	C_1	Times between which $y^{-\frac{1}{2}}=b_1t+C_1$ (seconds)	y_0*	$\log_{10}B^\delta$	ρ
20	100	0·535	0·303	0·14–0·55	51·8	$\bar{1}$·505	19·9
	68·8	0·553	0·386	0·16–0·39	49·0	$\bar{1}$·605	19·8
	35·3	0·780	0·402	0·13–0·35	28·5	$\bar{1}$·616	20·6
	25·3	0·925	0·444	0·12–0·39	16·7	$\bar{1}$·672	19·3
	17·8	0·950	0·496	0·11–0·32	11·9	$\bar{1}$·697	20·4
	11·0	1·24	0·754	0·11–0·40	4·25	$\bar{1}$·873	20·1
	9·5	1·54	1·067	0·14–0·40	2·35	0·033	19·8
	7·6	1·58	1·124	0·10–0·28	2·00	0·047	21·8
40	100	0·780	0·256	0·09–0·34	—	$\bar{1}$·4405	17·7
	68·8	0·880	0·436	0·19–0·40	28·5	$\bar{1}$·6585	17·3
	35·3	0·915	0·538	0·14–0·31	15·0	$\bar{1}$·7460	17·5
60	100	0·980	0·308	0·15–0·30	—	$\bar{1}$·514	12·1
	68·8	2·49	0·388	0·08–0·20	21·0	$\bar{1}$·6255	11·4
	35·3	2·54	0·496	0·09–0·18	15·0	$\bar{1}$·720	11·0
80	100	2·58	0·246	0·03–0·16	—	—	—
	68·8	4·58	0·265	0·02–0·18	—	—	—
	55·3	4·72	0·434	0·03–0·14	—	—	

* y_0 is the value of y at $t=0$.

(5.16) can now be solved for ζ. This is easy if more manageable expressions for C are first found: they are obtained by substituting expansions for S_u in powers of $\delta \equiv 1/(\rho - 1)$ (Appendix A) in (5.8), viz.

$$\beta_1 n_{10} t + C = \frac{1}{u}[1 + \delta \log_e (1 + \zeta u^{\rho-1}) + 0(\delta^2 \zeta u^{\rho-1})] \text{ if } \zeta u^{\rho-1} < 1 \qquad \ldots\ldots(5.17)$$

$$= \frac{\pi\delta\zeta^{\delta}}{\sin \pi\delta} + \frac{1}{u}\left[\delta \log\left(1 + \frac{u^{1-}}{\zeta}\right) - 0\left(\delta^2 \frac{u^{1-\rho}}{\zeta}\right)\right] \text{ if } \zeta u^{\rho-1} > 1. \qquad \ldots\ldots(5.18)$$

Then putting $t=0$ (i.e. $u=1$) immediately yields C.* It is easily seen that the ζ values corresponding to $A_F = \frac{1}{2}$ will differ from those corresponding to $A_F = 2$. It is therefore necessary to decide which value to take for A_F. The ratio of the contribution to the luminescence of centres of the second kind to that of the first kind at $t=0$ is $\rho\zeta$, which (from the approximate values of the ζ obtained by putting $C=0$) tends to increase with exciting intensity for both values of A_F and, for $I=100$, $T=20$° c., $=1{\cdot}4$ if $A_F=2$ and $7{\cdot}0$ if $A^F = \frac{1}{2}$. Now the greater part of the fluorescent emission (figure 2) measured at intensities somewhat higher than those of our observations is due to silver centres, and so it seems reasonable to identify centres of the second kind with silver and those of the first kind with zinc,† i.e. $A_F = 2$.

The ζ being known, (5.17) with (5.6) also gives values of X and t over all the decay period. It was used in particular to calculate the times (corresponding to $u^{\rho} = \frac{1}{2}$) taken for the number of positive holes of the centre with the faster decay (silver) to fall to half its initial value. The first term suffices for calculating the much longer half-value times (corresponding to $u = \frac{1}{2}$) for the zinc centres. These times are tabulated with other derived constants of physical interest, including the ratio of the observed to calculated values of y_0, in table 2 *b*.

The constants (tables 2 *a* and 2 *b*) vary with I in an unexpectedly complicated way. At 20° c., $\beta_1 n_{10}$ appears to vary periodically with I: its values at 40, 60° and 80° c. are not inconsistent with a similar behaviour. Moreover, almost all other constants as functions of I appear to have periodic terms superimposed on a simpler function. If these periodicity effects are real, they might be explained if the differential equations governing the build-up have periodic terms: in Appendix D we show that this may be so under certain conditions.

There are, however, several features of the results which cannot be explained. Thus the ratio of the estimates by two different methods of quantities proportional to the total number of excited electrons from blue centres are different for each set of observations at 20° c. The first method of estimation is given in the build-up section, where it is shown that the total number of excited electrons is proportional to the product of the exciting intensity (I) and the time of activation (t_0). The second method depends on the fact that the total number of electrons excited from blue centres is $\int X\,dt$ integrated over build-up and decay. The value of this integral over the decay could be obtained by

---

* It is interesting to note that when δ is small, $C \simeq 1$ even if ζ is large. This means that the final decay equation is changed only slightly by the presence of a second centre with a considerably faster β even if the number of its holes is large.

† The more so as rough measurements of the spectral distribution indicate that the ratio of the contribution from zinc and silver centres decreases with exciting intensity.

Table 2 *b*. Derived constants

Temperature (°c.)	I	ζ	C	$\beta_1 A$	$\beta_1 n_{10}$	$\frac{\beta_1 n_{10}^2}{A}$	$\frac{n_{10}}{A}$	$\frac{n_{20}}{A}$	$\frac{\text{Observed } y_0}{\text{Calculated } y_0}$	Times (sec.)	
										$u=\frac{1}{2}$	$\frac{n_2}{n_{20}}=\frac{1}{2}$*
20	100·0	0·072	1·0039	0·286	1·773	10·99	6·20	0·447	1·134	0·5618	0·0198
$\rho=20$	68·8	0·058	32	0·306	1·437	6·76	4·70	0·273	2·063	0·6937	0·0244
	35·3	0·043	24	0·608	1·945	6·22	3·20	0·137	1·616	0·5123	0·0180
	25·3	0·077	40	0·856	2·084	5·08	2·44	0·189	0·748	0·4779	0·0169
	17·8	0·027	15	0·902	1·919	4·08	2·13	0·0574	1·366	0·5243	0·0183
	11·0	0·021	12	1·538	1·647	1·765	1·072	0·0225	1·282	0·6665	0·0213
	9·5	0·031	17	2·372	1·445	0·881	0·610	0·0189	1·155	0·6908	0·0244
	7·6	0·033	18	2·496	1·408	0·794	0·564	0·0186	1·052	0·7089	0·0250
40	100·0	0·105	1·0065	0·608	3·067	15·47	5·04	0·529	—	0·3240	0·0121
$\rho=17\cdot5$	68·8	0·061	38	0·774	2·027	5·31	2·62	0·160	1·614	0·4914	0·0190
	35·3	0·082	50	0·837	1·710	3·49	2·04	0·167	1·027	0·5818	0·0222
60	100·0	0·084	1·0084	0·960	3·209	10·73	3·34	0·281	—	0·3090	0·0181
$\rho=11\cdot5$	68·8	0·113	113	6·20	6·490	6·80	1·048	0·118	0·771	0·1524	0·0087
	35·3	0·082	82	6·45	5·168	4·14	0·801	0·0657	1·161	0·1918	0·0113
80	100·0	—	—	6·80	10·5	16·55	1·58	—	—	0·0952	—
	68·8	—	—	21·0	17·3	14·24	0·826	—	—	0·0378	—
	35·3	—	—	23·3	10·9	5·32	0·489	—	—	0·0918	—

* $\frac{n_2}{n_{20}}=\frac{1}{2}$ corresponds to $u=0\cdot9658$ at 20° c., 0·9612 at 40° c. and 0·9415 at 60° c.

extrapolation from the last recorded observations (table 1); its values over the build-up could only be found for the 20° C. observations as the build-up results at the other temperatures were not complete. The values of the ratio (table 1) indicate that the decay must be slower than a -2 power law at long times. It is significant that measurements of the phosphorescent intensity of ZnS at long times published by Johnson (1939), Fonda (1945), Randall and Wilkins (1945 c) and Jesty (1946) do not satisfy the law of (5.11). Another unexplained feature is that the observed y_0 values do not agree with the calculated ones. Again, as Garlick and Wilkins (1945) have pointed out, the β should be proportional to the square root of the absolute temperature, but they vary with temperature more rapidly than that. Finally, the most serious difficulty is that the β should not depend on I, but they are, in fact, roughly proportional to $1/I$.

In the next section we shall show how these difficulties may be accounted for and the constants reinterpreted by allowing for the presence of trapping centres.

Two other factors have been neglected in developing the theory. Thus the phosphor layer is thick enough to cause considerable variations in the conditions at different depths. However, when β_1 decreases as n_{10} increases, so that $\beta_1 n_{10}$ does not vary greatly either with I or with depth in the layer, the law still holds. It is valid in particular for the amount of variation of $\beta_1 n_{10}$ at 20° C. in table 2 *b*. These statements are amplified in Appendix B. The effect of the copper centres has also been neglected. If the number of activated green centres per unit volume is n_3, we should put

$$\frac{dn_3}{dt}=\beta_3 n n_3, \qquad \ldots\ldots(5.19)$$

$$n=n_1+n_2+n_3. \qquad \ldots\ldots(5.20)$$

Equations (5.1) and (5.2) are unchanged except that n is given by (5.20) and (5.4) is unaltered since the green luminescence is filtered out. A justification for neglecting n_3 is suggested in the next section.

§6. PROPOSED THEORY FOR PHOSPHORS HAVING MORE THAN ONE TYPE OF ACTIVATING CENTRE AND TRAPPING CENTRES

The phenomenon of thermoluminescence is well known. There is general agreement that it is due to electrons being trapped after excitation in some regions of the crystal (trapping or metastable levels) from which they can only be released by receiving thermal energy from vibrations of the lattice ions. Randall and Wilkins (1945 a, b) have studied the subject extensively: for example, they excited a phosphor for some time at low temperature and measured the phosphorescent emission when it was heated at a constant rate; they consider that the intensity-temperature relationship (glow curve) so obtained approximately corresponds to the trap depth distribution of the phosphor. The physical picture of the trapping levels is not very satisfactory; they have been associated with vacant negative ion lattice points, cracks and regions of strain in the crystal and surface levels.

The simplest assumption is that of de Groot (1939) and Blokhinzev (1937), who suggested that trapped electrons are released by absorbing thermal energy at a rate γl proportional to the number of electrons in traps, and that traps capture electrons at a rate $\alpha n(L-l)$ proportional to the concentration n of free electrons and the concentration of empty traps $L-l$, where the number of positive holes per unit volume is, of course, $n+l$. It is shown in Appendix C (iv) that for the two-centre case these assumptions lead at long times to the relations (5.1) and (5.2) but with the β_1 multipled by $\gamma/(\alpha L+\gamma)$. This ratio is the limiting value (R) of the ratio (r) of free electrons to holes.

Now γ is of the form $se^{-E/\boldsymbol{k}T}$, where E is the trap depth, T the absolute temperature and $\boldsymbol{k}$ is Boltzmann's constant, and according to Garlick and Wilkins (1945) β_1 and β_2 increase as the square root of T. Klasens and Wise (1946) have pointed out that if T is high or the traps are very shallow, $\gamma/(\alpha L+\gamma)\simeq 1$ and $\beta_1 R$ and $\beta_2 R$ increase only slowly with temperature, while if $\alpha L \gg \gamma$, electrons when excited spend most of the time in traps and $\beta_1 R$ and $\beta_2 R$ will increase more rapidly with temperature. However, there seems no reason why the ratio of the βR should change with temperature. Also it seems quite impossible for $\beta_1 R$ and $\beta_2 R$ to change with I.

We shall now show that these difficulties may be resolved by postulating another trapping mechanism. Since a change in I produces a change in the crystal only at the activating centres, we believe that there must be traps closely associated with the excited centres. We suggest, therefore, that when a positive hole is formed on an activating centre, shallow secondary potential holes which can trap electrons are produced in its neighbourhood: these secondary holes disappear when the primary positive holes are filled. Electrons can be released from the secondary holes by absorbing vibrational energy from neighbouring ions. An electron so released may either fall into the primary hole or wander through the lattice. The process is somewhat analogous to that of an electron digging its own hole (Landau (1933), Hippel (1936), Gurney and Mott (1937) and Mott (1937)).

It was assumed that there were $\lambda_j n_j (j=1, 2, 3)$ traps per unit volume associated with positive holes of the jth kind and that $\gamma_j l_j$ electrons vibrated out of them per second, of which $\gamma_j l_j p_j$ went straight into positive holes and $\gamma_j l_j(1-p_j)$ into the conduction band. It seemed likely that the rate at which free electrons fell into positive holes would depend on whether or not the associated traps were empty. We therefore considered in some detail the behaviour of n_j and l_j as functions of time when there was one trap per hole and the hole could not be filled by a free electron when the associated trap was occupied (Appendix C (i)). Of the various types of solution there was one which gave bimolecular laws at long times of the form of (5.1) and (5.2) but with β_j replaced by $\beta_j+\alpha_j p_j$. The proportion of excited electrons in traps could then be small even though free electrons were more often captured by traps than by holes. In the more general case, without the above restrictions, it was still found that β_j was replaced by $\beta_j+\lambda_j\alpha_j p_j$*: the result is independent of γ_j if γ_j is large and the traps are shallow, and physically this is equivalent to saying that traps delay the recombination.

It is known from observation that the decay of the green phosphorescence is much slower than that of the blue. Hence β_3 must be small compared with β_1 and, therefore, from the above theory either α_3 or p_3 must be small compared with α_1 and p_1 respectively. If p_3 is small, the traps associated with copper centres will behave like shallow de Groot-type traps with respect to the blue luminescence during the period of our observations. So their effect is to multiply the β_3 for the other centres by $\gamma_3/(\lambda_3\alpha_3 n_{30}+\gamma_3)$, where n_{30} is the value of n_3 at short times regarded as constant. Now n_{30} will be approximately proportional to I, and hence the observed increase of $1/\beta_1$ and $1/\beta_2$ with I is immediately explained.

---

* Cf. equations (C.24) and (E.3) on pp. 497 and 500 respectively.

The result of the new trapping theory is, therefore, that the quantities tabulated as $\beta_1 A$ and $\beta_2 A$ in table 2 *b* should be tabulated as $\beta_1' AR$, $\beta_2' AR$ or

$$(\beta_1+\lambda_1\alpha_1 p_1)A \bigg/ \left(1+\frac{\lambda_3\alpha_3 n_{30}}{\gamma_3}\right), \quad (\beta_2+\lambda_2\alpha_2 p_2)A \bigg/ \left(1+\frac{\lambda_3\alpha_3 n_{30}}{\gamma_3}\right) \text{ respectively.}$$

The ratio ρ is now β_2'/β_1' and is still independent of I. It is difficult to predict how it should change with T since the depth of the secondary potential holes as well as the energy distribution of the trapped electrons will depend on T. However, it is unlikely* that p_1 and p_2 will be invariant with T.

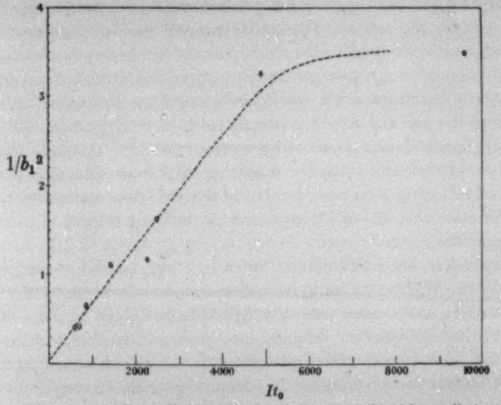


Figure 5. The variation of $1/b_1^2$ with intensity of exciting radiation.

There are enough values of $\beta_1' AR$ at 20° c. to justify examining the change with I more fully. From figure 5 it will be seen that apart from the periodic term and the point corresponding to $I100$, the values do lie about a straight line which nearly passes through the origin. This indicates that $\lambda_3\alpha_3 n_{30} \gg \gamma_3$ since $1/b_1^2=[1+\lambda_3\alpha_3 n_{30}/\gamma_3]/\beta_1' A$, and so most of the electrons are in traps associated with copper centres. This agrees with Garlick and Wilkins' (1945) conclusion that there are relatively few free electrons after the first few milliseconds.† It may mean that the copper traps are more numerous than the others or that their cross-section is larger or that they are deeper.

---

* See also hole migration discussion (p. 490).

† Klasens and Wise (1946) showed that with one activating centre and $\alpha=\beta$

$$r-R=(r_0-R)e^{-(\beta L+\gamma)t},$$

where $r=r_0$ at $t=0$. Using de Groot's (1939) estimate of $\beta L+\gamma$ (*c.* 100), $r-R$ is already small after a time of the order of 20 milliseconds.

Since, from the ratio of their concentrations, the number of copper centres is approximately 1/100 of the number of silver ones (§ 3), it follows that either there must be very many traps associated with each activated copper centre or that the ratio of excited to unexcited centres is much greater for copper than for the other centres. This can happen if either the absorption coefficient for copper centres is larger than for the others or if the numbers of excited centres are approaching their equilibrium values, when the ratio of excited Cu to excited Ag centres is $\mu_3\beta_2'/\mu_2\beta_3'$ (Appendix E) which can be large since $\beta_2' \gg \beta_3'$.

It is also possible that positive holes are migrating through the filled band to unactivated centres (Klasens, 1946), although the process is infrequent at low temperatures. An obvious effect would be that trapped electrons would escape to the conduction band, so that $\gamma_3(1-p_3)$ would increase more rapidly with temperature when hole migration was taking place. Another effect would be to increase the proportion of excited copper centres. This is because the probability that a hole will leave a blue centre is greater than the probability of one leaving a green centre, since the green energy levels are further from the filled band than the blue ones. The second process is small compared with the first if the cross-sections of hole capture for the different centres are of the same order, since the number of copper centres is $<1\%$ of the number of other centres. Further evidence for this is that the equilibrium blue fluorescent intensity is proportional to the exciting intensity (see figure 6, § 7): Klasens has shown that hole migration from blue to green centres causes the blue intensity to increase more rapidly with exciting intensity than this. It is also possible that a net migration from silver to zinc centres (whose energy levels are nearer the filled band than the silver ones are—§ 4*) may account for the tendency of ζ to increase with exciting intensity.

According to Randall and Wilkins (1945 b) the long-period phosphorescence is due entirely to electrons escaping from deep traps. In our case, the number so trapped must be of the same order as the number of green centres—if it were much greater, there would be a stronger long-period blue afterglow; if it were much less, the short-period green phosphorescence would be stronger. The number of electrons escaping from the deep traps will be negligible during the period of our observations, and hence their effect is to reduce the number of free electrons from $n_1+n_2+n_3-l$ to approximately n_1+n_2-l, thus justifying our neglect of the electrons from copper centres in the expression (5.3) for the number of free electrons in § 5.

The above work is based on the assumption that the traps associated with one type of centre are all of the same depth and cross-section. However, in view of the work of Randall and Wilkins (1945 a) this does not seem likely. We have therefore endeavoured to find the effect on our theory of assuming a continuous distribution of trap depths to be associated with each kind of centre.

For the zinc and silver centres, no general results were obtained, but it seemed likely (Appendix C (iii)) that, provided the final law was a power law, $\beta_1'(=\beta_1+\lambda_1\alpha_1p_1)$ generalizes to $\beta_1+\frac{\lambda_1}{n_1}\int\alpha_E p_E n_E dE$ integrated over all trap depths, where

---

* This migration from silver, of course, rapidly becomes negligible after the end of the excitation.

$n_E dE$ is the number of holes which have $\lambda_1 n_E dE$ traps of depths in the range E to $E+dE$ associated with them. α_E, γ_E and p_E refer to traps of depths E to $E+dE$ and have their usual significance. For the copper centres, regarding the associated traps as de Groot-type ones, it is found (Appendix C (v)) that if eventually the relationship between phosphorescent intensity and time can be represented by a power law, the power is always -2, and $L\alpha/\gamma$ is replaced by $\int \frac{\alpha_E L_E}{\gamma_E} dt$ integrated over all depths; all the earlier conclusions are therefore unaffected.

It also follows that a limiting power law cannot be due to traps unassociated with excited centres when the index is not -2. This is interesting in view of Randall and Wilkins' (1945 b) results on long-period decays. They found limiting power laws other than -2 experimentally in several cases: e.g. for a ZnS with silver and copper impurities, X was proportional to $t^{-1\cdot 32}$ for $t \simeq 20$ sec. $-$ 200 min. They showed theoretically that limiting power laws of any value < -1 are to be expected if the distribution of trap depths is exponential and it is assumed that the long-period phosphorescent intensity is proportional to the rate of excape of electrons from traps. They also found that the glow curve for phosphors with limiting power laws < -1 were approximately exponential in the region mainly responsible for the long-period phosphorescence, and that the limiting power laws calculated on the assumption that these glow curves give the trap-depth distribution agreed with the measured ones in several cases. Now this theoretical interpretation of the phosphorescence in terms of trap-depth distributions is only valid if an electron enters a hole almost immediately after it escapes from a deep trap and is never retrapped. This follows naturally from our theory if electrons go straight to copper holes from associated traps (i.e. $p_{3E} \simeq 1$ when E is large). On the other hand it seems unlikely that traps of such depth are produced merely by the ionization of Cu centres. Possibly the traps are due to localized lattice distortions associated with every copper centre whether activated or not, but can only be filled by electrons that have just been excited from the copper centres. This would explain also why the number of electrons in deep traps is about equal to the number of excited copper centres.

§7. THE BUILD-UP OF THE LUMINESCENCE

The equations for the build-up of the luminescence are difficult to discuss even when no trapping centres are involved because the effect of no kind of centre can be neglected.* So far we have only worked on the build-up equations corresponding to C.2 and C.3 (p. 495) with one trap associated with each hole.

If we assume that the number of positive holes of the jth kind is small compared with the number of such centres, then the rate at which electrons are ejected from centres of this type is $\frac{I\mu_j}{h\nu} (= k_j)$, where $\frac{I}{h\nu}$ is the number of exciting photons incident on this phosphor per second and μ_j is the probability that a particular photon excites any activating centre. Making the further assumption that the number of electrons falling straight back into their centres is negligible, and writing q_j for the probability that an ejected electron wanders through the crystal as a conduction electron, then $1-q_j$ must be the probability that it falls immediately into the associated trap during that period. The differential equations governing the build-up of fluorescence are then

---

* For this reason we have not included the build-up observations. At 20° c. they showed that the period of activation was never long enough for the fluorescence to reach its equilibrium value.

$$\frac{dn_j}{dt}=k_jq_j-(\beta_j+\alpha_j)nn_j+\gamma_j(1-p_j)l_j, \qquad \ldots\ldots(7.1)$$

$$\frac{dl_j}{dt}=k_j(1-q_j)+\alpha_jnn_j-\gamma_jl_j, \qquad \ldots\ldots(7.2)$$

where $j=1$, 2, or 3,

$$X=k_1+k_2-\frac{d}{dt}(n_1+n_2+l_1+l_2). \qquad \ldots\ldots.(7.3)$$

The linearity of the relationship between the equilibrium fluorescent and exciting intensities (illustrated in figure 6) is consistent with these equations since in conjunction with the condition for the equilibrium of the fluorescent intensity, viz.

$$\frac{dn_j}{dt}=0=\frac{dl_j}{dt},$$

they obviously lead to the relationship $X_{0j}=k_j$ where X_{0j} is the equilibrium fluorescent intensity. Thus it is justifiable to neglect the number of photons encountering activated centres.

It is shown in Appendix D that the solutions of these equations can have no periodic term when the asymptotic values of n_j and l_j are very nearly reached, but may do so during the preceding stages of the build-up. Since the solution for n_2 in terms of n_1 ((7.1) ($j=2$)) contains n_1 integrated once, the periodic parts of the expressions for n_2 and n_1 differ in phase by a right angle. Approximately the same phase difference would be expected between the curves of n_{10} and n_{20} versus exciting intensity values corrected to the same activation time, since increasing I roughly corresponds to speeding up the build-up. The number of our observations is not, however, large enough to permit of making this test.

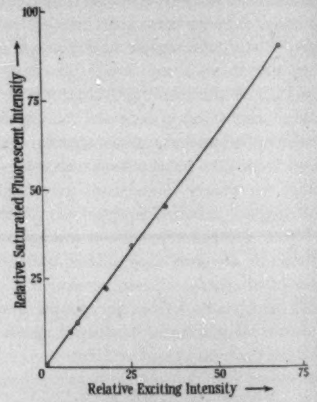


Figure 6. The variation of the saturated fluorescent intensity with intensity of exciting radiation.

If the presence of trapping centres is neglected, it is difficult to tell whether or not the solutions of the build-up equations have any periodic term. They can, however, have a maximum (Appendix E): such an effect has been observed by de Groot (1939 b). This unexpected result is given in full in Appendix E, where the equilibrium values of n_1 and n_2 (after long excitation) are also derived with and without traps.

APPENDIX A

Derivation of the series used in the two-centre decay analysis

Equation (5.8) gives

$$\beta_1 n_{10} t + C = \frac{1}{u+\zeta u^{\rho}} + \zeta(\rho-1) S_u, \qquad \ldots\ldots(\text{A.1})$$

where $$S_u = \int_0^u \frac{u^{\rho-3}\,du}{(1+\zeta u^{\rho-1})^2} \text{ and } t=0 \text{ when } u=1.$$

In the integrand $(1+\zeta u^{\rho-1})^{-2}$ is expanded in powers of $\zeta u^{\rho-1}$ or $u^{1-\rho}/\zeta$ whichever <1. In the latter case, which is the more difficult, we put

$$-S_u + \int_0^{\infty} \frac{u^{\rho-3}\,du}{(1+\zeta u^{\rho-1})^2} = \int_u^{\infty} \frac{u^{\rho-3}\,du}{(1+\zeta u^{\rho-1})^2}. \qquad \ldots\ldots(\text{A.2})$$

In the integral on the left we put $\zeta u^{\rho-1} = y^2$. The resulting integral is obtained in terms of a contour integral round an infinite rectangle in the upper half-plane. Integrating the right-hand side term by term after expanding,

$$\begin{aligned} S_u &= 2\zeta^{\delta-1}\delta \int_0^{\infty} \frac{y^{1-2\delta}\,dy}{(1+y^2)^2} - \frac{1}{\zeta^2}\left\{\frac{u^{-\rho}}{\rho} - \frac{u^{1-2\rho}}{\zeta(\rho-\frac{1}{2})} + \frac{u^{2-3\rho}}{\zeta^2(\rho-\frac{2}{3})} - \frac{u^{3-4\rho}}{\zeta^3(\rho-\frac{3}{4})} + \ldots\right\} \\ &= \zeta^{\delta-1}\frac{\pi\delta}{\sin\pi\delta} - \frac{1}{uS(\rho-1)}\left\{\frac{u^{1-\rho}}{\zeta\left(1+\frac{1}{\rho-1}\right)} - \left(\frac{u^{1-\rho}}{\zeta}\right)^2 \frac{1}{1+\frac{2}{\rho-1}}\right. \\ &\qquad \left. + \left(\frac{u^{1-}}{\zeta}\right)^3 \frac{1}{1+\frac{3}{\rho-1}} \cdots\right\}. \qquad \ldots\ldots(\text{A.3}) \end{aligned}$$

Expanding the denominators, writing $\delta = 1/(\rho-1)$, collecting powers of δ and substituting in (A.1),

$$\beta_1 n_{10} t + C = \frac{\zeta^{\delta}\pi\delta}{\sin\pi\delta} + \frac{1}{u}\left\{\delta \log_e\left(1+\frac{u^{1-\rho}}{\zeta}\right) - \delta^2 z\left(2, \frac{u^{1-\rho}}{\zeta}\right) + \delta^3 z\left(3, \frac{u^{1-\rho}}{\zeta}\right)\right\}, \qquad \ldots\ldots(\text{A.4})$$

where $$z(g, x) = x - \frac{x^2}{2^g} + \frac{x^3}{3^g} - \frac{x^4}{4^g} \qquad \ldots\ldots(\text{A.5})$$

By a very similar process we obtain the expansion of S_u when $\zeta u^{\rho-1} < 1$:

$$\beta_1 n_{10} t + C = \frac{1}{u}\left\{1 + \delta \log_e(1+\zeta u^{\rho-1}) + \delta^2 z(2, \zeta u^{\rho-1}) + \sum_{g=3}^{\infty} \delta^g z(g, \zeta u^{\rho-1})\right\}. \qquad \ldots\ldots(\text{A.6})$$

z has been tabulated for all values of g when $x=1$ (Jahnke and Emde, 1943). The values for $g = 2, 3, 4$ are respectively ·8225, ·9015, ·9470. We have calculated $z(2, x)$ for some other values of x:—

Table 3. Values of $z(2, x)$ and $\frac{z(2, x)}{x}$

x	0	0·2	0·4	0·6	0·8	1·0
$z(2, x)$	0	0·1908	0·3658	0·5281	0·6798	0·8225
$\frac{z(2, x)}{x}$	1	0·9540	0·9146	0·8802	0·8497	0·8225

With a fairly small δ graphical interpolation of $z(2, x)$ is adequate in all cases. and for the terms involving larger g's we can replace z/ζ by approximations independent of ζ, viz:

$$\sum_{g=3}^{\infty}(-1)^g\delta^g z(g,x) \doteqdot \frac{\delta^3 x}{1+\delta} \doteqdot x\left(0{\cdot}9\delta^3+\frac{\delta^4}{1+\delta}\right), \qquad \ldots\ldots(\text{A.7})$$

and

$$\sum_{g=3}^{\infty}\delta^g z(g,x) \doteqdot \frac{\delta^3 x}{1-\delta} \doteqdot x\left(0{\cdot}9\delta^3+\frac{\delta^4}{1-\delta}\right) \qquad \ldots\ldots(\text{A.8})$$

APPENDIX B

Decay laws when n is not constant throughout the layer

As mentioned in § 5, we have assumed that n does not vary from place to place in the powder at a particular instant. 60% of the crystals are less than 8 microns in diameter, which is less than 10 times de Groot's (1940) estimation of the mean path of electrons in the conduction band; since they move freely, we would expect equations (5.1) to (5.4) to be exact for single crystals in the layer, but we must allow for observing luminescence from many crystals at once, at different depths.

About 99% of ultra-violet light was absorbed and about 75% of white light. There is much scattering, so we would expect this absorption to be higher than for a single crystal. De Groot estimated that it is 10^3 cm.$^{-1}$ at 3700 A. for a single crystal. If d_0 is the thickness and λ the coefficient, we take

$$e^{-\lambda d_0} \doteqdot {\cdot}01 \quad \text{or} \quad \lambda d_0 = 4{\cdot}605. \qquad \ldots\ldots(\text{B.1})$$

Taking $d_0 = 0{\cdot}008$ cm. this gives $\lambda = 577$ cm.$^{-1}$ in our case. However, d_0 is probably over-estimated and the void should also reduce the absorption coefficient for a packed powder below that for one crystal.

Jesty (1946) summed intensities for a number (up to 20) of thin layers of material.

De Groot (1940) calculated the effect of summing intensity-time relationships from crystals at continuously varying depths, taking the simplest possible function of t for X, viz.,

$$X = \frac{\beta_1 n_{10}^2}{(1+\beta_1 n_{10} t)^2}. \qquad \ldots\ldots(\text{B.2})$$

n_{10} is assumed to vary continuously. De Groot (1940) assumed that n_{10} varied exponentially with depth. We think this is the best approximation for a packed layer of fine particles. The law is unaltered for short and for long times; at intermediate values a different law is produced. But in no case does $\frac{d}{dt}(X^{-\frac{1}{2}})$ vary with the initial value of n_1, and hence with the initial exciting intensity, so that the change of β with I cannot be explained in this way.

If, however, β, as is our case, varies with I, it must also vary with depth in the layer; if the surface is activated for a time t_0 with intensity I_0, the corresponding intensity at a depth d is $I_0 e^{-\lambda x}$, where $x = d/d_0$. We should also allow for the fact that some of the phosphorescent light is absorbed in its way through the layer. The contribution to X at a depth d should therefore be multiplied by $e^{-\lambda_b(1-x)}$, where $e^{-\lambda_b}$ is the proportion of blue light transmitted by the whole layer, and

$x=d/d_0$, d_0 being the total depth. In our case this gives $\lambda_b \doteqdot 2\log_e 2$. Using suffixes x to denote values of $\beta_1 n_{10}^2$ and $\beta_1 n_{10}$ at depth x, the complete single-centre approximation is

$$X=\int_0^1 \frac{(\beta_1 n_{10}{}^2)_x e^{-\lambda_b(1-x)}}{(C_x+(\beta_1 n_{10})_x t)^2}\left[1+\frac{1+\rho+\dfrac{2}{\rho-2}}{(C_x+(\beta_1 n_{10})_x t)^{\rho-1}}\right]dx. \qquad \ldots\ldots\text{(B.3)}$$

Since the values of $\beta_1 n_{10}$ and $\beta_1 n_{10}{}^2/A$ deduced from observations correspond to (table 1) about the same periods of activation, we can estimate from them the approximate amount that $(\beta_1 n_{10})_x$ and $(\beta_1 n_{10})_x/A$ vary for a fixed I. $\beta_1 n_{10}$ (table 3, column 9) varies by about ±20% at 20° c. over a wide range of values of I, and if $(\beta_1 n_{10})_x$ varies by this amount in (B.3) it is found that a law of the form (B.2) is a very close aproximation to (B.3) even if $\beta_1 n_{10}^2$ varies much more.

Approximate empirical relationships, obtained from table 3, are

$$\beta_1 n_{10}=1\cdot 75+0\cdot 35\cos 2\pi(\log I-1\cdot 4)=f_1(I), \qquad \ldots\ldots\text{(B.4)}$$

$$\log_e \frac{\beta_1 n_1{}^2}{A}=2\log_e I-1\cdot 5+0\cdot 47\cos(\log_{10} I-1\cdot 45)=f_2(I). \qquad \ldots\ldots\text{(B.5)}$$

In estimating these quantities as a function of the depth, we assumed that we could put

$$(\beta_1 n_{10})_x=f_1(\theta I e^{-\lambda x}), \qquad \ldots\ldots\text{(B.6)}$$

and

$$\log_e\left(\frac{\beta_1 n_{10}^2}{A}\right)_x=f_2(\theta I e^{-\lambda x}), \qquad \ldots\ldots\text{(B.7)}$$

where θ is a constant fairly near to 1 that is the same for all curves at 20° c. It is equivalent to assuming that at a fixed depth $\bar{x}\left(=\frac{\log\theta}{\lambda}\right)$, $(\beta_1 n_{10})_{\bar{x}}$ and $\left(\frac{\beta_1 n_{10}^2}{A}\right)_{\bar{x}}$ have the values of (B.4) and (B.5) for a given I. This is obviously accurate if the main contribution to x comes from a thin slice of the layer near to the surface, which is so if $\beta_1 n_{10}^2$ increases rapidly with I. The effect needs to be investigated further, especially with more than one activating centre present.

APPENDIX C

Decay equations for ionized activating centres and trapping centres

1. It is first assumed that one trap is associated with each positive hole, which cannot be filled from the conduction band when the trap is occupied by an electron.

l_1 is the concentration of electrons trapped at centres of type 1, n_1 the number per unit volume of positive holes associated with unoccupied traps, n the concentration of free electrons. Then

$$X=-\frac{d}{dt}(n_1+l_1+n_2+l_2), \qquad \ldots\ldots\text{(C.1)}$$

$$n=n_1+n_2, \qquad \ldots\ldots\text{(C.1')}$$

$$\frac{dn_j}{dt}=-(\alpha_j+\beta_j)nn_j+\gamma_j l_j(1-p_j), \qquad \ldots\ldots\text{(C.2)}$$

$$\frac{dl_j}{dt}=\alpha_j nn_j-\gamma l_j. \qquad \ldots\ldots\text{(C.3)}$$

Various solutions have been obtained. In each case l_j in terms of n_j is given by

$$l_j=\left\{l_{j0}+\alpha_j\int_0^t nn_j e^{\gamma_j t}\right\}e^{-\gamma_j t}. \qquad \text{......(C.4)}$$

(i) Putting $n=n_j+N_j$ and substituting for l_j,

$$\frac{d^2n_j}{dt^2}+\frac{dn_j}{dt}\{\gamma_j+(\alpha_j+\beta_j)N_j+2(\alpha_j+\beta_j)n_j\}+n_j\left\{(\alpha_j+p_j)\frac{dN_j}{dt}+(\alpha_j p+\beta_j)\gamma_j N_j\right\}$$
$$+n_j^2\gamma_j(\alpha_j p_j+\beta_j)=0. \qquad \text{......(C.5)}$$

If N is changing so slowly that it can be regarded as constant, the solution is the sum of two exponentials when second-degree terms in N and N_j are negligible, viz.,

$$n_j=A_1e^{-\lambda_1 t}+A_2e^{-\lambda_2 t},$$

where λ_1 and λ_2 are given by

$$2\lambda=\gamma_j+(\alpha_j+\beta_j)N_j\pm\sqrt{\{\gamma_j-(\alpha_j+\beta_j)N_j\}^2+4\gamma_j\alpha_j(1-p_j)N_j}. \qquad \text{......(C.6)}$$

(ii) When N is negligible and n_1 is large there is an expansion valid at short times:

$$n_1{}^2=y^2\left\{1+\frac{A_1}{y}+\frac{A_1{}^3(A_2+2)}{24A_2y^3}+\dots\right\}, \qquad \text{......(C.7)}$$

where

$$A_1=\frac{\gamma_1\{\beta_1+\alpha_1(2-p_1)\}}{2(\alpha_1+\beta_1)^2},\qquad A_2=\frac{\beta_1+(2-p_1)\alpha_1}{\alpha_1p_1+\beta_1},\qquad y=ae^{\frac{\frac{1}{2}\gamma_1(\alpha_1p_1+\beta_1)}{\alpha_1+\beta_1}}$$

and a is a constant.

With this expansion l_1 can be nearly proportional to $n_1{}^2$ and there is an approximate bimolecular law, but under limited conditions and for a short period.

(iii) At long times there is an exponential series satisfying (C.5) with $N_j=0$ viz.:

$$n_j=x+\tfrac{1}{2}(B_1-B_2)x^2+\frac{x^3}{3!}(\tfrac{3}{2}B_1-B_2)(B_1-B_2)\dots,$$

where $$x=a_1e^{-\gamma_j t},\quad B_1=\frac{2(\alpha_j+\beta_j)}{\gamma_j},\quad B_2=\frac{\alpha_jp_j+\beta_j}{\gamma_j} \qquad \text{......(C.8)}$$

so l_1 tends to vary linearly with n_1.

(iv) The equations are also satisfied when n_1 is a negative power of t at long times. If nn_j is a sum of negative powers of t, so that $\frac{d}{dt}(nn_j)$ is small compared with nn_j,

$$\frac{l_j}{nn_j}\sim\frac{\alpha_j}{\gamma_j}. \qquad \text{......(C.9)}$$

Hence $$\frac{dn_j}{dt}\sim-(\beta_j+\alpha_jp_j)nn_j, \qquad \text{......(C.10)}$$

$$\frac{dl_j}{dt}\sim 0. \qquad \text{......(C.11)}$$

2. Case (iv) was next considered with λ_j traps per positive hole of the jth kind when it was assumed that free electrons fell into holes at a rate that did not depend on the number of associated trapped electrons.

If there are l_j trapped electrons and n_j holes per unit volume,

$$\frac{dn_j}{dt} = -\beta_j n_j n - \gamma_j p_j l_j, \qquad \text{......(C.21)}$$

$$\frac{dl_j}{dt} = \alpha_j n(\lambda_j n_j - l_j) - \gamma_j l_j, \qquad \text{......(C.22)}$$

The equation for n_j in terms of n and t, eliminating l_j, was much more complicated than (C.5), but it was found that if n and n_j were sums of negative powers of t, then

$$l_j \sim \frac{\lambda_j p_j \alpha_j}{\gamma_j} n n_j, \qquad \text{......(C.23)}$$

$$\frac{dn_j}{dt} \sim -(\beta_j + \lambda_j \alpha_j p_j) n_j^2. \qquad \text{......(C.24)}$$

3. We next consider the effect of variations of depth, cross-section and "p" in traps associated with positive holes.

Let there be $n_E dE$ holes with associated traps of depths between E and $E + dE$ of which there are $\lambda_1 n_E dE$, and $l_E dE$ trapped electrons.

Then the number of holes is

$$n_1 = \int_0^\infty n_E dE, \qquad \text{......(C.31)}$$

$$\frac{dn_E}{dt} = -\beta_1 n n_E - \gamma_E p_E l_E, \qquad \text{......(C.32)}$$

$$\frac{dl_E}{dt} = \alpha_E n(\lambda_1 n_E - l_E) - \gamma_E l_E, \qquad \text{......(C.33)}$$

$$\left.\begin{aligned} l_1 &= \int_0^\infty l_E dE, \\ n &= n_1 - l_1, \\ X &= -\frac{dn_1}{dt}. \end{aligned}\right\} \qquad \text{......(C.34)}$$

At times when $l^{-\gamma_E t}$ and $\frac{1}{n n_E \gamma_E} \frac{d}{dt} (n n_E)$ are negligible these equations are satisfied by

$$l_E = \frac{\lambda_1 \alpha_E}{\gamma_E} n_E n. \qquad \text{......(C.35)}$$

Hence

$$\frac{dn_E}{dt} = -(\beta_1 + \lambda_1 \alpha_E p_E) n_E n \qquad \text{......(C.36)}$$

and

$$\frac{dn_1}{dt} = -\beta_1 n_1 n - \lambda_1 n \int \alpha_E n_E p_E dE. \qquad \text{......(C.37)}$$

4. *Traps independent of centres (de Groot type) of one depth.* With two kinds of activating centres the number of free electrons is

$$n = n_1 + n_2 - l, \qquad \text{......(C.41)}$$

$$\frac{dl}{dt}=\alpha n(L-l)-\gamma l, \qquad \ldots\ldots(\mathrm{C}.42)$$

$$\frac{dn_1}{dt}=-\beta_1 n_1 n, \qquad \ldots\ldots(\mathrm{C}.43)$$

$$\frac{dn_2}{dt}=-\beta_2 n_2 n. \qquad \ldots\ldots(\mathrm{C}.44)$$

Substituting for l from (C.43) in (C.42) leads to

$$\frac{d^2n_1}{dt^2}-\frac{1}{n_1}\left(\frac{dn_1}{dt}\right)^2\left(1+\frac{\alpha_1}{\beta_1}\right)+\frac{1}{n_1}\frac{dn_1}{dt}\{\beta_1 n_1+\beta_2 n_2-\alpha(n_1+n_2)+\underline{\gamma+\alpha L}\}$$
$$+\underline{\beta_1\gamma(n_1+n_2)}=0, \qquad \ldots\ldots(\mathrm{C}.45)$$

and there is an analogous equation with suffixes 2 and 1 interchanged.

It is easily verified that the terms underlined can ultimately become the largest. (5.1) and (5.2) are then obtained with $\beta_1 R$, $\beta_2 R$, in place of β_1 and β_2. $R=\dfrac{1}{1+\alpha L/\gamma}$, and the remaining terms are of the order of $1/t^2$.

5. *De Groot type traps of different depths.* Let there be $L_E dE$ traps with depths between E and $E+dE$, with $l_E dE$ electrons trapped and escaping at a rate $\gamma_E l_E dE$. Then

$$\frac{dl_E}{dt}=\alpha_E n(L_E-l_E)-\gamma_E l_E. \qquad \ldots\ldots(\mathrm{C}.51)$$

(C.43) is unchanged: substituting from it for n,

$$\frac{dl_E}{dt}+l_E\left\{\gamma_E-\frac{\alpha_E}{\beta_1 n_1}\frac{dn_1}{dt}\right\}=-\frac{\alpha_E L_E}{\beta_1 n_1}\frac{dn_1}{dt}. \qquad \ldots\ldots(\mathrm{C}.52)$$

Then if $l_E=l_{E0}$ at $t=0$,

$$l_E=e^{-\gamma_E t}n_1^{\frac{-\alpha_E}{\beta}}\left\{l_{E0}n_{10}^{\frac{-\alpha_E}{\beta}}+L_E\int_0^t e^{\gamma_E t}\frac{d}{dt}\left(n_1^{\frac{-\alpha_E}{\beta}}\right)dt\right\}. \qquad \ldots\ldots(\mathrm{C}.53)$$

By integrations by parts,

$$\frac{\gamma_E}{L_E}\left\{l_E e^{\gamma_E t}n_1^{\frac{-\alpha_E}{\beta}}-l_{E0}n_{10}^{\frac{-\alpha_E}{\beta}}\right\}=e^{\gamma_E t}\left\{\frac{d}{dt}\left(n_1^{\frac{-\alpha_E}{\beta}}\right)-\frac{1}{\gamma_E}\frac{d^2}{dt^2}\left(n_1^{\frac{-\alpha_E}{\beta}}\right)\right\}$$
$$-\frac{d}{dt}\left(n_1^{\frac{-\alpha_E}{\beta}}\right)+\frac{1}{\gamma_E}\frac{d^2}{dt^2}\left(n_1^{\frac{-\alpha_E}{\beta}}\right)+\frac{1}{\gamma_E}\int_0^t e^{\gamma_E t}\frac{d^3}{dt^3}\left(n_1^{\frac{-\alpha_E}{\beta}}\right)dt. \qquad \ldots\ldots(\mathrm{C}.54)$$

So if the intensity $\dfrac{-dn_1}{dt}$ is a negative power of t, when $e^{-\gamma_E t}$ and $\dfrac{1}{\gamma_E t}$ are negligible,

$$l_E\sim\frac{\alpha_E L_E}{\gamma_E\beta_1 n_1}\frac{dn_1}{dt}, \qquad \ldots\ldots(\mathrm{C}.55)$$

or

$$\frac{l_E}{n}\sim\frac{\alpha_E L_E}{\gamma_E}. \qquad \ldots\ldots(\mathrm{C}.56)$$

Thus the ratio of concentrations of trapped electrons to free electrons tends to $\int\dfrac{\alpha_E L_E\,dE}{\gamma_E}$ integrated over the whole distribution of depths, whether the numbers of free electrons per unit volume is n_1-l, n_1+n_2-l, or $n_1+n_2+n_3-l-l_\infty$ (l_∞ being the number in deep traps).

The most general equations satisfied by our observations are, therefore,

$$\frac{dn_j}{dt} = -\frac{\beta_j' n_j n}{1+\lambda_3 \int \frac{\alpha_3 n_{3E}\, dE}{\gamma_{3E}}},$$

where $$\int n_{3E}\, dE = n_3 \doteqdot n_{30} \quad \text{and} \quad j=1 \text{ or } 2. \qquad \ldots\ldots(\text{C. } 57)$$

APPENDIX D

Conditions for a periodic term in the build-up equation with more than one type of activating centre and associated trapping centres

We shall use the notation of equations (C.5) and remove the suffixes, which will all be alike. n is now the number per unit volume of ionized activating centres of the jth kind and $n+N$ is the number of electrons in the conduction band. Then eliminating l_j between (7.1) and (7.2) with this notation gives

$$\frac{d^2n}{dt^2} + \frac{dn}{dt}\{\gamma + (\alpha+\beta)(N+2n)\} + n\left\{(\alpha+\beta)\left(\gamma N + \frac{dN}{dt}\right) - \gamma\alpha N(1-p)\right\}$$
$$+ \gamma n^2(\alpha p + \beta) = k\gamma(1-p+pq).$$

Let $n = y_0 + y_1$, where y_0 satisfies (D.1) when N is constant. If we put $N = N_0 + \tau$, where τ is a small function of t, and neglect second-degree terms in y_1 and τ, y_1 satisfies

$$\frac{d^2y_1}{dt^2} + \frac{dy_1}{dt}\{\gamma + (\alpha+\beta)(N_0+\tau+2y_0)\} + y_1(\alpha p+\beta)\gamma(N_0+\tau+2y_0)$$
$$+ (\alpha+\beta)\left(2\frac{dy_0}{dt} + \frac{d\tau}{dt}\right) = -y_0\left\{(\alpha+\beta)\frac{d\tau}{dt} + \tau(\alpha p+\beta)\right\}. \qquad \ldots\ldots(\text{D.2})$$

The same equation could hold during the decay, but with a different y_0. It may be reduced to a standard form by a process explained in many text-books (e.g. Levy, *Numerical Studies in Differential Equations*, §§18, 18.1). If the equation is

$$\frac{d^2y}{dt^2} + P\frac{dy}{dt} + Qy = R,$$

where P, Q and R are functions of t only, we put $y = uv$, where $u = e^{-\frac{1}{2}\int P\,dt}$, so that

$$\frac{d^2v}{dt^2} + vD(t) = R,$$

where $$D(t) = Q - \tfrac{1}{2}\frac{dP}{dt} - \tfrac{1}{4}P^2. \qquad \ldots\ldots(\text{D.3})$$

Solutions are generally in terms of solutions with $R=0$. We can then compare (D.3) with corresponding equations with $D(t)$ constant, and may expect a periodic type of solution when $D(t)$ is positive and not when it is negative. For equation (D.2),

$$D(t) = \tfrac{1}{2}(\alpha+\beta)\left\{2\frac{dy_0}{dt} + \frac{d\tau}{dl}\right\} - \tfrac{1}{4}\{\gamma - (\alpha+\beta)(N_0+\tau+2y_0)\}^2$$
$$- y\alpha(1-p)(N_0+\tau+2y_0). \qquad \ldots\ldots(\text{D.4})$$

During the build-up $D(t)$ will not be positive when the asymptotic values of l and n are nearly reached, but may be so at earlier stages, when dy_0/dt and $d\tau/dt$

must be positive most of the time. During most of the decay it is almost certainly negative.

APPENDIX E

Approximations to the build-up equations after long times of activation.

The equations are

$$\frac{dn_1}{dt} = k_1 - \beta_1 n_1(n_1+n_2), \qquad \text{......(E.1)}$$

$$\frac{dn_2}{dt} = k_2 - \beta_2 n_2(n_1+n_2). \qquad \text{......(E.2)}$$

The equilibrium values are easily found by putting

$$\frac{dn_1}{dt} = \frac{dn_2}{dt} = 0.$$

In the most general case, from equations (7.1), (7.2),

$$\operatorname*{Lt}_{t\to\infty} n_j = M_j = \frac{K_j}{\alpha_j p_j + \beta_j}\left\{\sum_i \frac{K_i}{\alpha_i p_i + \beta_i}\right\}^{-\frac{1}{2}}, \qquad \text{......(E.3)}$$

where
$$K_j = k_j(1 - p_j + p_j q_j).$$

With no trapping centres,

$$M_1 = \frac{k_1}{\beta_1}\left(\frac{k_1}{\beta_1} + \frac{k_2}{\beta_2}\right)^{-\frac{1}{2}}, \quad M_2 = \frac{k_2}{\beta_2}\left(\frac{k_1}{\beta_1} + \frac{k_2}{\beta_2}\right)^{-\frac{1}{2}}. \qquad \text{......(E.4)}$$

We can now write

$$\left.\begin{aligned} n_1 &= M_1 - a_{11}e^{-\lambda_1 t} - a_{21}e^{-2\lambda_1 t} - a_{31}e^{-3\lambda_1 t}, \\ n_2 &= M_2 - a_{12}e^{-\lambda_1 t} - a_{22}e^{-2\lambda_1 t} - a_{32}e^{-3\lambda_1 t}. \end{aligned}\right\}. \qquad \text{......(E.5)}$$

Equating coefficients of $e^{-\lambda_1 t}$ gives the equations

$$a_{11}\{\lambda_1 - \beta_1(2M_1+M_2)\} = a_{12}\beta_1 M_1, \qquad \text{......(E.6)}$$

$$a_{11}\beta_2 M_2 = a_{12}\{\lambda_1 - \beta_2(M_1+2M_2)\}. \qquad \text{......(E.7)}$$

Eliminating a_{11} and a_{12}, λ_1 must be the smallest root of

$$\lambda_1^2 - \lambda_1\{\beta_1(M_2+2M_1) + \beta_2(M_1+2M_2)\} + 2\beta_1\beta_2(M_1+M_2)^2 = 0. \qquad \text{.....(E.8)}$$

The roots are positive and real, and we can write

$$\begin{aligned} 2\lambda_1 &= \beta_1(2M_1+M_2) + \beta_2(M_1+2M_2) \\ &\quad - \sqrt{M_1^2(2\beta_1-\beta_2)^2 + M_2^2(\beta_1-2\beta_2)^2 + 2M_1M_2(2\beta_1^2 - 3\beta_1\beta_2 + 2\beta_2^2)}. \end{aligned} \qquad \text{......(E.9)}$$

If λ_2 is the other root, the solutions will be sums of exponentials when indices are linear combinations of λ_1 and λ_2. For large enough values of t all but the first exponential terms will be negligible, so that

$$n_1 \doteqdot M_1 - a_{11}e^{-\lambda_1 t}, \qquad \text{......(E.10)}$$

$$n_2 \doteqdot M_2 - a_{12}e^{-\lambda_1 t}. \qquad \text{......(E.11)}$$

From (E.6) and (E.4)

$$\begin{aligned} \frac{a_{11}}{a_{12}}\beta_2 M_2 &= \tfrac{1}{2}\beta_1 M_1\left(2 + \frac{k_2\beta_1}{k_1\beta_2} - \frac{\beta_2}{\beta_1} - \frac{2k_2}{k_1}\right) \\ &\quad - \tfrac{1}{2}\sqrt{M_1^2(2\beta_1-\beta_2)^2 + M_2^2(\beta_1-2\beta_2)^2 + 2M_1M_2(2\beta_1^2 - 3\beta_1\beta_2 + 2\beta_2^2)}, \end{aligned} \qquad \text{......(E.12)}$$

which is negative if $\beta_2 > 2\beta_1$, and probably in other cases also.

Thus either n_1 or n_2 rises above its equilibrium value. If $\beta_2 \gg \beta_1$ and k_1 and k_2 are of the same order of magnitude, the solution for n_1 will not differ greatly from that with n_2 negligible, so n_2 must then have a maximum.

§ 8. CONCLUSION

The emphasis in this work has been on the detailed investigation of the time variation of the luminescence over a limited period rather than a general investigation over long periods. Even so, the analysis of the observations assuming a bimolecular law and two kinds of activating centre indicates that for quantitative work on the first few milliseconds of the decay yet more detailed observations are required during that period.

The idea that a phosphor may contain more than one kind of activating centre is not new (Martin and Headrick (1939)), but as far as is known no work has previously been done on analysing decay curves into components due to the various centres when electrons are excited into the conduction band. For our phosphor, during fluorescence and the very early stages of the decay, the contribution from the silver centres is most important; during the intermediate stage of the decay, the contribution from zinc centres predominates and at long times the emission is mainly due to copper centres. However, each centre has some effect at all times.

The variations with temperature and exciting intensity of the constants obtained by the two-centre analysis have led to a new conception of the trapping mechanism and to the conclusions that most of the electrons are in traps associated with copper centres and that during the period of our observations retrapping is important. We have also shown that these conclusions do not depend on the nature of the trap-depth distribution and that the presence of deep traps of other kinds is not excluded.

The number of our decay curves is not sufficient to justify quantitative work on the variation of the constants. It is interesting to note that variations in conditions at different depths in the phosphor are more or less compensated for in our case by variations with I of the constants. This probably does not happen in general, and so for thick layers of other phosphors the decay laws may be very complicated (e.g. Jesty (1946)).

It is thus seen that the nature of the time variation of the phosphorescence depends on many factors. We have neglected the fact that our observations correspond to intermittent excitation (ratio excitation : decay periods : : 1 : 18) and have only mentioned hole migration. In studying the time variation of the fluorescence there are even more factors to take into account, and so little has been done on this.

Finally, it can be said that although we modify the simple bimolecular theory considerably, it remains the basis of our interpretation.

ACKNOWLEDGMENTS

The authors are indebted to Mr. R. S. Longhurst for assistance with the calculations and to Mr. W. Ramsden for measuring the spectral distribution. They also thank Dr. H. A. Klasens for stimulating criticisms and suggestions, and Mr. J. A. M. van Moll and the directors of Philips' Lamps Ltd. for permission to publish this paper.

REFERENCES

Beese, 1939. *J. Opt. Soc. Amer.*, **29**, 26.
Blokhinzev, 1937. *Phys. Z. Soviet Union*, **12**, 586.
Fonda, 1945. *Trans. Electrochem. Soc.* Preprint 87–9.
Garlick and Wilkins, 1945. *Proc. Roy. Soc.*, A, **184**, 408.
de Groot, 1939 a. *Physica*, **6**, 275.
de Groot, 1939 b. *Physica*, **6**, 393.
de Groot, 1940. *Physica*, **7**, 432.
Gudden and Pohl, 1921. *Z. Phys.*, **4**, 206.
Gurney and Mott, 1937. *Proc. Phys. Soc.*, **49** (extra part), 32.
Henderson, 1939. *Proc. Roy. Soc.*, A, **173**, 323.
v. Hippel, 1936. *Z. Phys.*, **101**, 680.
Jahnke and Ende, 1943. *Tables of Higher Mathematical Functions*, p. 323.
Jesty, 1946. Lecture delivered at the Research Laboratories of the General Electrical Company, 15 June
Johnson, 1939. *J. Opt. Soc. Amer.*, **29**, 387.
Kitchener, 1939. *Trans. Faraday Soc.*, **35**, 97.
Klasens, 1946. *Nature, Lond.*, **158**, 306.
Klasens and Wise, 1946. *Nature, Lond.*, **158**, 483.
Landau, 1933. *Phys. Z. Soviet Union*, **3**, 664.
Leverenz and Seitz, 1939. *J. Appl. Phys.*, **10**, 479.
Lewschin and Romanowsky, 1934. *Phys. Z. Soviet Union*, **5**, 379.
Martin and Headrick, 1939. *J. Appl. Phys.*, **10**, 116.
Milner, 1939. *Trans. Faraday Soc.*, **35**, 101.
Mott, 1937. *Nature, Lond.*, **139**, 951.
Randall and Wilkins, 1945 a. *Proc. Roy. Soc.*, A, **184**, 365.
Randall and Wilkins, 1945 b. *Proc. Roy. Soc.*, A, **184**, 390.
Rees, 1942. *Ann. Rep. Chem. Soc.*, **39**, 76.
Riehl, 1939. *Trans. Faraday Soc.*, **35**, 135.
Seitz, 1938. *J. Chem. Phys.*, **6**, 454.
Seitz, 1939 a. *Trans. Faraday Soc.*, **35**, 83.
Sietz, 1939 b. *Trans. Faraday Soc.*, **35**, 98.

ADDENDUM TO DISCUSSION

on the paper by R. F. Schmid and L. Gerö entitled "Photochemical decomposition of CO" (*Proc. Phys. Soc.*, **58**, 701 (1946)).

Dr. J. G. Valatin*. The comments by Schmid and Gerö on the interpretation of the photochemical decomposition of CO seem to be at variance with the views of Gaydon, but they do not disagree with available direct experimental evidence on the absorption spectrum. As is emphasized by Schmid and Gerö, they conclude from the observed predissociation effect on the $A^1\Pi$ state that the effective absorption region of the continuum at 77497 cm.$^{-1}$ is very narrow, and the transition probability is small ; it can give rise to a considerable effect in the case of a suitably chosen and concentrated, monochromatic light source, but the effectiveness of the continuum cannot be judged from plates taken with low resolving power. The spectrograms of Leifson can give no evidence at all on the effect in question in the region of the xenon line.

The three lines of evidence given in the paper of Schmid and Gerö show that the absorption of the 1295 A. xenon line cannot be due to the lines of the Fourth Positive band system. Quite apart from the discussion of an eventual larger overall width of these lines, the lines of the Fourth Positive band system belong in the corresponding spectral region to initial states of the CO molecule which are not present at ordinary temperature.

---

* Since both authors are now dead, Dr. Valatin has asked if he may reply to the questions raised by Dr. Gaydon in the discussion which is printed immediately after the paper.

As to the dissociation energy of CO, the discussion still holds. (*See also* Long and Norrish, 1946 ; Valatin, 1946 ; Edlén, 1947.) The interpretation of the photochemical decomposition of CO gives further support to the dissociation scheme of Schmid and Gerő. The value of 170 kcal./gm.-atom deduced for the heat of sublimation of carbon into ^5S atomic states agrees well with the dynamic experiments, while the lower value of the equilibrium measurements can be explained by primary formation of ^5S atoms and by secondary processes. Corresponding calculations, resulting also in vapour-pressure curves which are in good agreement with the equilibrium measurements in the carbon arc, have been given by Schmid and Gerő (1937). Papers dealing with the thermochemical side of the question are in preparation.

REFERENCES

Edlén, B., 1947. *Nature, Lond.*, **159**, 129.

Long, L. H., and Norrish, R. G. W., 1946. *Nature, Lond.*, **157**, 486 ; **158**, 237.

Schmid, R., and Gerő, L., 1937. *Mitteilungen der berg- und hüttenmännischen Abteilung der Universitat, Sopron, IX.*, 173.

Valatin, J. G., 1946. *Nature, Lond.*, **158**, 237.

OBITUARY NOTICES

Sir JAMES JEANS, O.M., F.R.S.

Sir James Jeans, famous alike as mathematician, astronomer and mathematical physicist, and world-famous as an expositor of all three sciences, died at his home, Cleveland Lodge, Dorking, on 16 September 1946 of coronary thrombosis. He had had heart attacks a year or so previously, and had had to reduce his activities, but otherwise he had been in reasonably good health ; his last few hours were passed in intense pain.

Jeans made outstanding contributions to theoretical physics on the one hand and to astronomy and cosmogony on the other hand. In each field of thought he solved some of the most difficult problems of the day. But he was not only gifted as an investigator ; he was also superbly gifted as a writer. Besides his text-books, and his two treatises, *The Dynamical Theory of Gases* and *Astronomy and Cosmogony*, he wrote two masterpieces : the one, his Physical Society Report of 1914, *Report on Radiation and the Quantum Theory*, a gem of economical exposition which ranks with Eddington's Physical Society Report of 1918, *Report on the Relativity Theory of Gravitation*, as having substantially influenced the general acceptance of a new and fundamental physical theory ; the other, his Adams Prize Essay of 1917, published in 1919 under the title *Problems of Cosmogony and Stellar Dynamics*, which unfolds in thrilling style, but with full mathematical detail, the classical researches on the stability of forms of equilibrium of rotating masses to which Jeans himself made the dominant original contributions. These works would have sufficed for any ordinary man. But in 1928, almost suddenly, feeling perhaps that his best original work was finished, Jeans turned from research to popular exposition, and at once attained a justly-deserved success. In a series of volumes, showing no signs of the speed with which they must have been composed, Jeans traversed the ground of his own and others' researches in astronomy and physics, covering nebular, stellar and planetary evolution, thermodynamics, atomic theory, relativity and quantum theory in a fresh and engaging style, illustrating the varying orders of magnitude of astronomical and atomic quantities with many a vivid simile. Whether it were technical mathematics, detailed account of an original theory, popular astronomy, popular physics or popular philosophy, Jeans hardly wrote a dull sentence. And I am tempted to put alongside his two technical masterpieces, his semi-popular volume *Science and Music* (1938), wherein he showed a side of himself which had previously been developed only in his private life. For the great mathematician, daring speculator, modern physicist and (to be truthful) interested, but only adequate, philosopher, that Jeans in turn was, was also a lover of music, a performer on the organ, a builder of two organs at his own home and a music-room designer who thought it worth while, at the age of 60, to combine his scientific and musical knowledge for the benefit of still another public. His distinction at whatever he touched equalled his versatility. And though Jeans formed no school of research in the ordinary sense, the world of science is most emphatically the poorer by his unexpected loss.

Yet Jeans had his limitations, and to get the best out of his writings it is necessary to appreciate what these were. His limitations were those of a mathematician (I speak as a mathematician) who likes the actual problems of the universe and of the atom formulated in a tidy way, with their ragged ends all tucked in, so that it is possible to make general statements about them with possibly undue confidence that they are true. It was small wonder that in his Rede lecture, *The Mysterious Universe* (1928), he considered the Great Architect of the Universe to be a Mathematician ; that anthropomorphic and mechanical models of the universe all failing, the only elements of reality Jeans could associate with the spectacle of Nature consisted of pure thought, " the thought of one whom we must consider, for want of a wider word, as a mathematical thinker ". It is part of the same characteristic that Jeans as a physicist was happiest in dealing with the *general* problems of physics, the state of molecular chaos in a gas, the nature of the second law of thermodynamics, the equipartition of energy and the distribution of energy between matter and radiation, and that he was less successful in dealing with actual stars, interiors or exteriors, as physical objects. No one was more sure than Jeans in his grasp of physical *principles* ; but there was not the same reality about the material of which his stars were made as there was about Eddington's. The time is not yet for a comparison of these two Titans. But it may be permissible to draw attention to the fact that Eddington, originally an astronomer, and only later a physicist, had a deeper physical insight than Jeans, whilst Jeans, the mathematician, had legitimate grounds for criticizing the mathematical processes by which Eddington appeared to get some of his results. Hence their occasional clashes, and hence their widely differing conclusions about stellar constitution. But it is pleasant to be able to record that when the Gold Medal of the Royal Astronomical Society was awarded to Jeans in 1922 the presentation was made by Eddington, accompanied with one of the latter's characteristically eloquent addresses.

James Hopwood Jeans was born at Ormskirk, near Southport, on 11 September 1877, the son of William Tulloch Jeans, a parliamentary journalist ; he had two younger sisters. As a boy he was much interested in numbers ; he discovered his father's book of logarithms when he was 7 and, being unable to understand what they were for, learned the first 20 of them off by heart—and retained them in his memory till near the end of his life. He was also much interested in clocks, and wrote a short booklet, " Clocks, by J. Jeans ", at the age of 9. He went to Merchant Taylors' School from 1890 to 1896, and then entered Trinity College, Cambridge, and read mathematics. He was bracketed Second Wrangler in the Mathematical Tripos of 1898 and took a First Class in Part II of the same Tripos in 1900. He spent some time in the Cavendish Laboratory whilst holding an Isaac Newton studentship. He won a Smith's Prize in 1900 and was elected a Fellow of Trinity in 1901. About this time he had a spell of ill-health, suffering from tuberculosis of the joints, and had to spend some time in sanatoria, but he made a complete recovery. He became a University Lecturer in Mathematics at Cambridge in 1904, but from 1905 to 1909 he held a Chair of Applied Mathematics at Princeton. He returned to Cambridge as Stokes Lecturer in 1910 but resigned the post in 1912, thereafter holding no regular university appointment. He was elected Professor of Astronomy in the Royal Institution in 1935, and was annually re-elected until he resigned from ill-health in 1946.

Jeans was elected a Fellow of the Royal Society in 1906, at the early age of 28. He delivered the Bakerian Lecture, on " The Configurations of Rotating Compressible Masses ", in 1917, and was awarded that Society's Royal Medal in 1919. In the latter year he became an honorary secretary of the Royal Society, holding the post for the full period of 10 years.

He was created a Knight in 1928. He became a Research Associate of Mount Wilson Observatory in 1923. He was President of the Royal Astronomical Society for 1925–27, and President of the British Association at its Aberdeen meeting in 1934. He was given honorary degrees at a number of universities, at home and abroad. He became an Honorary Fellow of the Institute of Physics in 1929. He was awarded the Franklin Medal of the Franklin Institute in 1931. The supreme distinction of the Order of Merit was bestowed on him in 1939. He became an Honorary Fellow of Trinity in 1942.

In addition to the Bakerian Lecture of 1917, Jeans delivered, amongst others, the following formal lectures :—the Halley Lecture (1922) on " The Nebular Hypothesis and Modern Cosmogony ", the Guthrie Lecture of the Physical Society in 1923 on " The Present Position of the Radiation Problem (*Proc. Phys. Soc.*, **35**, pp. 222–224, 1923), the Rouse Ball Lecture (1925) on " Atomicity and Quanta ", the Rede Lecture (1930) on " The Mysterious Universe ", the Van der Waals Lecture (1923) on " The Physical Significance of Van der Waals' Equation ", the Kelvin Lecture (1925) on " Electric Forces and Quanta ", the H. H. Wills Lecture (1928) on " The Physics of the Universe, and the Silvanus Thompson Memorial Lecture (1931) on " What is Radiation ? "

Jeans married in 1907 Charlotte Tiffany Mitchell, daughter of Alfred Mitchell, of New London, Conn. She died in 1934, leaving one daughter. He married secondly, in 1935, Suzanne Hock, daughter of Oscar Hock, of Vienna and formerly Prague. There are two sons and a daughter of the second marriage. The first Lady Jeans gained some reputation as a poetess. The second Lady Jeans is a concert organist, and has given concerts on tours at home and abroad.

I have already mentioned Jeans' musical interests. He had an organ built for himself at his home at Dorking during the lifetime of the first Lady Jeans. When he married again, he had a second organ constructed for Lady Jeans, after an antique pattern which he called a " baroque " organ, and he had a special room built for his own organ, the two rooms being acoustically insulated from each other so that he and his wife could play without either disturbing the other. Jeans had played the organ from the age of 12, but he would never perform even before close friends. He could play the whole of Bach's organ works, and his preferences were for contrapuntal compositions (Lady Jeans played Bach at his funeral). His *Science and Music* covers an enormous range of physics, and it has been highly praised. Starting with the acoustics of the human ear, it went on to explain, for the benefit of the non-specialist, the nature of pure tones, of scales and keys, of the various musical instruments, and concluded with an account of the relation of an orchestra to a concert hall, the materials to use and the optimum size of a hall for a given orchestra. The writer remembers once falling in with Jeans at an " open " day at the N.P.L., when demonstrations in the acoustics department were being given, and being amazed at the wealth of technical knowledge about sound that Jeans had at his finger ends.

Jeans' scientific work was divided between physics and astronomy. In the latter field his main contribution was to the series of forms of equilibrium of rotating, gravitating, incompressible and compressible masses, and their stability. He finally settled the difficult problem of the stability of Poincaré's " pear-shaped " figure of a rotating liquid, showing how Sir George Darwin had been misled into considering it as stable. From the now-demonstrated instability, Jeans inferred a cataclysmic origin for double stars as produced by fission of rotating masses of stellar order ; he inferred that compressible masses would in general develop a lens-shaped figure with a sharp equatorial edge, from which matter would be ejected at two antipodal points determined by the tidal action of the rest of the universe, and he saw in this a possible origin for the forms of spiral nebulae. But in none of the effects of pure rotation could he find anything resembling the solar system, or system of one large body with much smaller bodies circulating round it. To account for the occurrence of the solar system he invoked the tidal effects of a passing star, which would raise jets in the primitive sun, these jets condensing into planets and yielding massive planets at middle distances, smaller ones further out and closer in. But he reckoned that such encounters would be excessively rare events. Jeans did much more for cosmogony than merely indulging in speculations. His Adams Prize Essay of 1919 contains the backbones of many fundamental calculations in this field, which must form the starting points of future investigations which will take into account the phenomenon of the expanding universe. Many of Jeans' specific conclusions will naturally need revision as time goes on. But *Problems of Cosmogony* is a great book. *Astronomy and Cosmogony* (1928) sums up all Jeans' original researches in cosmogony, but it carries less conviction than *Problems of Cosmogony*, and the physics in it, with its hypothesis that stars consist in parts of elements of atomic number 95 or more, is somewhat strained. It should be mentioned that this decade (1918–28) of Jeans' activities saw the publication of some 35 papers in the *Monthly Notices* of the Royal Astronomical Society, amongst which must especially be recorded his papers on " Radiative Viscosity ", in which he developed the cosmical importance of the transfer, by means of radiation, of angular momentum from one layer of a rotating body to another.

Perhaps astronomy was Jeans' favourite study; its attractions for him are seen in both his earliest and latest original works. But there was a period, say 1902–1914, when physical investigations claimed his principal attention. In paper after paper (chiefly in the *Philosophical Magazine*) he built up methods for developing the statistical mechanics of matter, and radiation in equilibrium with matter, for establishing on as rigorous a basis as possible Maxwell's distribution law for molecular velocities in a gas, and the theorem of equipartition of energy amongst the different degrees of freedom, and for determining the partition of energy between matter and radiation. These researches treated a gas of N molecules, where N is large, as a single dynamical system. He applied similar methods to the "ether" in an enclosure, resolving the fluctuating electromagnetic field therein into its harmonic constituents, calculating its number of degrees of freedom, and finally establishing what is now known as the Rayleigh-Jeans formula for the distribution of energy, in wave-length, in black-body radiation, on the classical theory, namely, $8\pi RT\lambda^{-4}d\lambda$. Rayleigh had previously published a similar formula with a different numerical factor, but at once accepted Jeans' form. This paper of Jeans of 1905 was a culminating point. He went on working at the enigma of radiation, attempting to find classical means whereby nature avoids the "ultra-violet catastrophe" predicted by the Rayleigh-Jeans formula. Eventually Jeans came to accept the law of radiation in the form given by Planck, though he suggested modifications in the mode of its derivation. He concluded that there must be something akin to a discontinuity in Nature's conduct of the process of the interchange of energy between matter and radiation. Jeans was no uncritical acceptor of the Quantum Theory; it was only after paper on paper, trying every conceivable resource to avoid the break with classical mechanics, that he was finally converted.

He then wrote his Physical Society *Report on Radiation and the Quantum Theory* (1914), in which he incorporated the then new Bohr theory of the hydrogen spectrum. This was a superb piece of exposition. He concluded it by remarking that "the keynote of the old mechanics was continuity, *natura non facit saltus*. The keynote of the new mechanics is discontinuity; in Poincaré's words: "Un système physique n'est susceptible que d'un nombre fini d'états distincts; il saute d'un de ces états à l'autre sans passer par une série continue d'états intermédiaires." And he ended by a free translation of a further passage from Poincaré's *Dernières Pensées*.

Though Jeans often lectured on the fundamentals of the quantum theory after his 1914 report, he made no distinctively original further contributions; his interests settled down to astronomy. But it is singularly appropriate that Jeans should have found Poincaré's words the most apt with which to close an epoch in his own life. For there is a very close parallel between the scientific interests of Poincaré and Jeans. Both were at once mathematicians, astronomers, physicists and philosophers; both wrote fundamental memoirs on the forms of equilibrium of rotating fluids; both devoted much thought to cosmogony; both were attracted by the early ideas on the quantum theory; and both wrote explicitly popular scientific books of high literary value. Both were scientific stylists. It is perhaps as an expositor that future generations will most cherish the memory of Jeans. Whether he were writing text-book, treatise, original paper or popular volume, he was always graphic, always fluent, always (or almost always) convincing, always a stimulus to the reader's curiosity. He came to stand for modern physics and astronomy to the people at large; and he richly deserved the full stature of that position. Physics and astronomy owe him much. But he was also an acute and tireless investigator, and it is as an investigator that his friends and colleagues will best care to remember him. E. A. MILNE.

THOMAS HOWELL LABY, M.A., Sc.D., F.R.S.

THE death of Prof. T. H. Laby, at the age of sixty-six, brings to an end the career of one devoted to the furtherance of physics and of science generally. His efforts in this direction, often in very difficult circumstances, undoubtedly undermined his health. Despite this, he was untiring in his work, the value of which to his native country, Australia, and to the world of science in general, is still inadequately recognized.

Laby was born in Victoria, Australia, and received his early academic education at the University of Sydney. After graduating, he was awarded an Exhibition of 1851 Overseas Research Studentship, and proceeded to Emmanuel College, Cambridge, and to research work at the Cavendish Laboratory under J. J. Thomson. After a successful period there, during which he held the Joule Studentship of the Royal Society, he took up an appointment as professor of physics at Wellington, New Zealand, in 1909. This he held until 1915, when he was elected to the chair of natural philosophy in the University of Melbourne, a post he retained until his resignation in 1942. His influence on Australian physics during this period was remarkable, and it is largely due to him that Australia holds a high place in the realm of physics.

Among the many reasons why Laby played such a unique part in the development of physics in Australia was his great interest in research and the wide range of his own activities in this direction. During his tenure of the chair, there existed throughout his department an air of enthusiasm and a feeling of complete confidence in the importance of the subject, which lent a distinction apparent to undergraduates as well as research students. This led to a remarkably regular production of very keen research students—so regular, in fact, that it was a matter of great surprise if, in any year, one of the Exhibition of 1851 Overseas Studentships did not fall to a member of Laby's department.

His primary interest was in precision experimental physics, but this did not prevent him from realizing the importance of other branches of the subject. Thus he was keenly aware of the importance of theoretical physics and encouraged any students with a bent in that direction. His unusual breadth of view is exhibited by his abolition of practical examinations in the subject, despite his own special interest in experiment.

It is difficult to say in which field of precise experiment Laby was most interested ; thermal conduction, mechanical equivalent of heat, x rays, geophysics, scientific radio, all occupied his attention and were a continual source of research problems for his students and assistants. The precision determination of J by Laby and Hercus is well known, as are also the series of papers by Laby and by his assistant Kannuluik on problems of thermal conduction. Laby was actively interested in the work of the geophysical prospecting party, led by Broughton Edge in Australia in 1929, and collaborated with Edge in editing the final report of the work, which is by way of becoming a standard text-book on the subject. Besides these researches, in which he, personally, took an active share, Laby encouraged work on nuclear physics, and a neutron generator was in operation just before the War.

Among his publications the most widely used is undoubtedly the ***Tables of Physical and Chemical Constants***, compiled in collaboration with Dr. G. W. C. Kaye, and now in its ninth edition.

Apart from his academic activities, Laby played a very important part in official developments in Australian science, such as the organization of the radium supply for hospitals and the formation and operation of the Radio Research Board. As a result of the latter, Australian workers have made, and are continuing to make, very important contributions to problems of radio transmission through the atmosphere. Despite all his other interests Laby maintained a detailed knowledge of developments in radio-physics. Thus, during his visit to England in 1934, he read to the Royal Society a stimulating paper by Martyn and Pulley, and was instrumental in exciting the interest of atomic physicists in ionospheric problems. He was thoroughly convinced of the importance of physics in the development of Australian industry, and devoted a great deal of time and effort towards the often thankless and wearisome task of convincing others of this now generally accepted fact.

At the outbreak of war in 1939, there existed virtually no optical industry in Australia to meet the requirements of optical munitions supply. Laby took a leading part in the organization of the Optical Munitions Panel of Australia, of which he was the first chairman. This body was vital to the establishment of a sufficiently productive industry. Laby's real value in the war crisis cannot be measured only by this. The great contribution that Australian physicists were able to make to the defence of their country and of the British Commonwealth could not have been made if in preceding years a firm tradition of high-quality physics had not been established in Australia, largely by the efforts of the Department of Natural Philosophy at Melbourne under Laby's direction. The difficulties of doing this under conditions of isolation imposed by the great distance of Australia from

Europe and America cannot easily be over-estimated, and there is no doubt that Laby sacrificed himself unsparingly in achieving this end.

H. S. W. MASSEY.

[Reprinted by permission from *Nature*, **158**, 157 (1946).]

GEORGE BLACKFORD BRYAN, O.B.E., D.Sc., M.I.E.E.

DR. G. B. BRYAN, formerly Professor of Physics, Royal Naval College, Greenwich, died at Nottingham on 29 November 1946 within a few days of his 72nd birthday. Educated at Nottingham High School and Nottingham University College, he took his B.Sc. (London) degree in 1896 with first-class honours. He gained many scholarships and prizes during this period, and from 1894 to 1896 undertook a research on electric waves on long wires in conjunction with the late Professor E. H. Barton. The results of this work were published in 1897 (*Proc. Phys. Soc.* and *Phil. Mag.*, January 1897).

These successes gained for him an 1851 Exhibition enabling him to enter St. John's College, Cambridge, for three years' research at the Cavendish Laboratory, working on the conductivity of thin liquid layers and on contact potentials under the direction of Sir J. J. Thomson. This brought him the B.A. degree by research and the D.Sc. (London).

Bryan's first appointment was as Demonstrator to the late Prof. A. M. Worthington, F.R.S., at the Royal Naval Engineering College, Devonport. In 1910 he came with Worthington to Greenwich, and on the latter's retirement in the following year he became the senior member of the staff under the Head of the Physics Department. In 1922 he was appointed to the Professorship, which he held until his retirement in 1938, in which year he was made an Officer of the British Empire in recognition of his great work for the Royal Navy. Bryan joined the Physical Society in 1916 and served on the Council from 1921 to 1926.

During the 1914–1918 war, working in co-operation with H.M. Signal School, Portsmouth, he built what was probably the first continuously evacuated triode valve and carried his investigations far enough for a 100-kw. valve to be planned with every hope of success. In 1940, and at an age when most men would have felt unable to take up new work, he responded to an urgent request to join the staff of the City and Guilds College as a Special Lecturer in order to assist in the intensive radio training of undergraduates entering the College under the Hankey scheme.

With his natural ability, Bryan's training and experience rendered him clever in devising and making apparatus for experimental work and for lecture demonstrations. His interests were mainly in the direction of applied electrical science, but he was also a successful interpreter of modern physical theories to young naval officers. He wrote and published little, not from lack of either ability or energy, but—as it seemed to his friends—from a feeling of diffidence. He did not seek publicity, and seemed to prefer serving those more in the limelight to attempting to establish a place for himself. He was a great tennis player and was good in other ball games.

One of the most loyal of colleagues, he had an exceptional power of making firm friends in all surroundings. He is sadly missed by all who knew him.

Bryan married Miss Ida Rodgers of Nottingham and he leaves one daughter.

C. L. FORTESCUE.

WILLIAM BARRON COUTTS, M.A. B.Sc.

WE regret to record the death, on 16 December 1946, at the Radcliffe Infirmary, Oxford, following an operation, of Professor W. B. Coutts, who served on the Staff of the Military College of Science from 1919 to the date of his death. Born at Kinghorn, Fifeshire, in 1885, he was educated at Edinburgh University and, after a period as a schoolmaster, was commissioned in 1915 to the R.G.A. (S.R.). He served at Gibraltar, and it was his work on the Rock which created his life-long interest in the problems of fire direction and control. In 1917 he was recalled to join the 35th Advanced Class and, after completing the course and obtaining the p.a.c.—a distinction of which he was very proud—he was appointed first Instructor and then Senior Lecturer in Range-finding at the College. In 1938 he was promoted to be Assistant Professor of Fire-control Instruments, a position which he held until his death. He was also a member of the Council of the Optical Society and for a period of five years was one of its Honorary Secretaries. He was for many years a regular contributor on optical instruments to the *Journal of Scientific Instruments*.

He was beloved by his colleagues and students, to whom he was universally known as "Willie", and his gift of dry Scotch humour enlivened many otherwise dreary conferences and lectures. His specialized knowledge of optical and fire-control instruments and his ability to teach the subject to technical officers will be a great loss.

REVIEWS OF BOOKS

Physics and Experience, by BERTRAND RUSSELL. Pp. 26. (The Henry Sidgwick Lecture, delivered at Newnham College, Cambridge, 10th November 1945. Cambridge: The University Press.) 1*s*. 6*d*. net.

In this lecture Lord Russell addresses himself, with his accustomed clarity, to the question: "Assuming physics to be broadly speaking true, can we know it to be true, and, if the answer is to be in the affirmative, does this involve knowledge of other truths besides those of physics?" He gives no clear-cut answer to this question, but concerns himself mainly with making clear its meaning and importance; only the first step towards an answer is taken at the end.

The portion of physics which is assumed to be true is that in which it is thought very unlikely that any new evidence will do more than somewhat modify it; for example, the wave theory of sound. This body of physics originated in percepts, but its constituents are very unlike the percepts which gave rise to them; our perception of sound does not at all resemble the wave. How, then, can we acquire a knowledge of the wave from the noise? Lord Russell holds the view that perception is the last link in an artificially limited portion of a causal chain of events which starts in a physical event and proceeds through physical space to the nerves and brain of the percipient; the percept "is what happens when, in common-sense terms, I see something or hear something or otherwise believe myself to become aware of something through my senses". This final event is not to be regarded as divisible into "perceiving" and an "object perceived"; it is a single unit. The percept which we call "seeing the Sun" is describable as a bright, hot, circular something existing at the moment of perception: the corresponding physical Sun is a spherical source of complex radiation existing eight minutes earlier. The former we know directly; the latter we infer from it. How is such an inference possible? Incidentally, Lord Russell includes both mental and physical events in the same causal chain, but distinguishes them by the fact that mental events alone can be known by someone otherwise than by inference. The physical world therefore becomes known only by inference—indeed, it almost follows, though he does not make the deduction, that it is *definable* as the world inferred from percepts—and the problem posited—how can it be so inferred?—stands out as a major problem of epistemology.

The one condition of true inference which Lord Russell allows himself to state is that the physical world must contain more or less separable causal chains; for if, for example, a chain started by the Sun interfered with one started by my neighbour's wireless set—or even with one started by the Moon—I should not be able, as I am, to infer the separate characteristics of those objects in the physical world. We know, however, that such independence is not complete, for when I look at sunlight reflected from a mirror I infer my face, but when I look at sunlight reflected from a tree I infer a tree. Hence the causal chains which both started in the Sun and both ended in my percepts are traced back to different arbitrary points because of the different degrees of interference of the reflecting surface in the two cases, and if I attempt to infer the character of the looking-glass from the appearance of my face, or that of the Sun from the appearance of the tree, my inference will be at least questionable. Hence "it is clear that the relation of a percept to the physical object which is supposed to be perceived is vague, approximate, and somewhat indefinite. There is no *precise* sense in which we can be said to perceive physical objects". One is left regretting that Lord Russell had no space to pursue the matter further because, obviously, in spite of this vagueness of relation between percept and corresponding physical object, there is, in fact, no uncertainty in my deduction about what I am seeing, and I experience no inconvenience through mistaking my face for a mirror or a tree for my face. He does add that " science consists largely of devices for overcoming this initial lack of precision on the assumption that perception gives a first approximation to the truth ", and one cannot help feeling that in the absence of an equally critical examination of those devices, the legitimacy of the preceding argument must be somewhat suspect. A line of thought which leads only to doubt of the truth of its premises must either be pursued to a final conclusion or exchanged for a more profitable one. We venture to suggest that the former alternative is impossible and the latter necessary.

What, in the most indubitable terms, can we say that the physicist does ? The answer, I think, is that he makes observations—acquires percepts—and then gives us a description of an external physical world which, if it behaved in a certain specified way, would produce those percepts. (To make discussion in a reasonable space possible I shall mean by " the physicist " someone to whom all physical observations generally accepted as trustworthy are his own percepts : we are thus not concerned with the problem of " other people "). The question then arises : should we say that the world which the physicist describes is an independently existing thing whose characteristics he " infers " from the percepts, or that it is a product of his creative imagination and formed so that it both entails the occurrence of the percepts and exhibits them as a correlated system ? There was a time when we should have said, without hesitation, that Lord Russell would choose the second mode of expression, for he has said somewhere that, as a general rule, one should speak in terms of construction rather than inference wherever possible. Here, however, he adopts the language of inference without so much as an apology, and so creates the problem with which he wrestles. He may, it is true, claim that the point in question is one on which it is not possible to speak in terms of construction, but this not only seems obviously false, but also invites the retort that it is the one question on which the rule has any importance. No one has any doubt about the appropriate language in limited considerations : Shakespeare constructed, not inferred, Caliban, and Einstein inferred, not constructed, the bending of light in a gravitational field. The difference is important only when we are considering the physical world as a whole, and if construction is not the legitimate concept there, then Lord Russell's advice loses its only important realm of application.

The question does not concern what the physicist does—on that, expressed in simple behaviouristic terms, there is probably general agreement—but rather what is the appropriate word for describing what he does. It is, nevertheless, not merely a verbal question, for if the physicist infers, the question arises, is what he infers true ? whereas if he constructs, the corresponding question is : does his construction serve his purpose ? The second question admits of an answer. It serves his purpose if it correlates his observations and does not entail anything contrary to observation ; and as new observations reveal defects in the contemporary construction he modifies it or destroys it and makes a better. This seems a faithful and adequate account of physical practice, and it is difficult to see why one need complicate it by adding the arbitrary postulate that the world-picture has a quality called " truth " or " untruth " which we must labour, without any hope of getting more than a probable answer, to determine.

The strong compulsion which many (not necessarily including Lord Russell) feel to making this addition presumably arises from the fact that in problems confined within a limited field of experience this additional quality of truth or untruth almost invariably signifies something necessary and important. If (as sometimes happens) I receive a letter stating that something which I have written is utterly absurd, I infer that the writer holds views different from mine. The inference is probably true, but it may be false ; he may, for instance, be a psychologist disinterestedly interested in observing my reaction, or one who is annoyed at having been convinced against his will. There is a meaning in saying that the inference is true or false, whether or not I am in a position to determine its truth or falsity, because independent tests are conceivable which would settle the matter ; and the possibility of those independent tests arises from the fact that experience is available outside the experience from which I made the inference. But when the question concerns the whole field of sense experience there is no independent source of information about the truth of the " inference ". We believe Lord Russell would hold that all that we can know about the physical world is what we can deduce from percepts. If, then, percepts exhaust their possibilities in giving us an " inference " without the " truth " label, we must necessarily remain in eternal ignorance whether that label can rightly be attached or not. In that case, what is the point of Lord Russell's problem ? It seems simpler not to bother about it.

There is another aspect of the question in which the distinction between partial and complete fields of experience is important. It is undoubtedly right to say that the physical world is brought into being in order to account for the very dissimilar world of percepts, but it is much less accurate to say that a particular element of the physical world is brought into being in order to account for a particular percept. Sound waves were not postulated to account for hearing, for, in fact, they do not account for it any better than a hypothesis of sound particles would do. Sound waves are part of a much larger body of hypotheses formed to account for a much larger field of experience, including, for example, the connection between the " velocity of sound " and the principal specific heats of the medium. If (as is by no means inconceivable since of the links in Lord Russell's causal chain from violin to percept the final one—from nerve disturbance to percept—is the one of which we know least, namely nothing) it should be found that the percept of hearing required sound particles rather than waves, we would not give up sound waves ; they would still be needed to account for the specific heat relation, and we should probably picture something like particles carried by waves, as was proposed in the somewhat analogous dilemma concerning light. The fact is that however smoothly the post-prandial narrative of the descent from vibrating string to sweet sound may trip off the tongue, it is the toil and sweat, blood and tears of the earlier climb that the gods exact from the philosopher, and he finds no passage from sound to string ; he must cut separate tracks from thermal, mechanical, visual, as well as auditory, percepts if he is to reach his goal.

The moral of all this seems to be that the construction or inference of the physical world from the world of percepts is such a complex matter, involving such an intricate network of connections, that the picking out of any single causal chain is highly artificial. Problems arising therefrom are arbitrary rather than inevitable. For instance, Lord Russell's original question : How can we infer from the world of percepts the very dissimilar world of physical objects ? is at least a plausible problem, but the question : How can we infer from the perceptual Sun the very dissimilar physical Sun ?, in so far as the dissimilarity of premises and conclusion is conceived as involving a difficulty, is not even that. At the first step, from percept to nerve disturbance, the resemblance is completely lost, and if there is a problem it is rather why the physical Sun ultimately arrived at should have recovered rather than lost so much resemblance to the percept. But it is hard to see what can be gained in any way by analysing the network into separate threads. All problems of physics are restricted to the physical world alone, percepts acting merely as a sort of Clerk of Works to ensure that the world is built according to specification, and all relevant problems of epistemology would seem to be concerned with the principles of inferring or constructing any kind of physical world from the whole assemblage of percepts, i.e. with the whole relation of specification to building, and not with the connection between each brick and its correlative clause. Notwithstanding Lord Russell's assumption that there is a considerable body of physics which will remain almost unchanged, we cannot neglect the possibility (I would even say probability), in view of the fate of gravitational force, light waves, eternal atoms and what not, that the present physical world will be completely transformed in the not very distant future. Our present causal chains will then lose their significance, but the general physical and epistemological problems will remain.

Physicists are not as a rule interested in the problems discussed here. Therein they are unfortunate, since they remain unaware of important implications of their achievements. Philosophers usually pay too little, but sometimes too much, respect to the work of physicists. Lord Russell is almost unique in maintaining a balanced judgment based on knowledge and understanding. The physicist who neglects what he has to offer loses much.

HERBERT DINGLE.

Methods of Mathematical Physics, by HAROLD JEFFREYS and BERTHA SWIRLES JEFFREYS. Pp. vii + 679. (Cambridge: The University Press, 1946.) *£3 3s. 0d.*

Although there have been rather a large number of books in the last few years with titles similar to this, it must not be dismissed as "just another of them." It is a book of quite exceptional importance, embodying as it does the considered ideas of one who has contributed very largely to mathematical physics. It is not specially intended for beginners, not is it a set of notes for the established worker. Rather it is a complete course, which, however, would only be intelligible to one who had had some previous mathematical training.

The first chapter is entitled the *real variable*, and takes up first the fundamental laws—associative, distributive etc.—of algebra, discusses whether the symbols in a physical equation like $s=ut$ represent numbers or physical magnitudes, including their units. Real numbers are next defined by the Dedekind *method of sections* and also by the *nest of intervals*. The remainder of the chapter is devoted to the properties of sequences, the theory of convergence (for the real variable), Riemann integration and the classical development of this theory, such as tests of convergence, mean-value theorems (including Taylor's series) and the like.

Chapter 2 gives the theory of vectors, including some applications to dynamics, and chapter 3 deals with tensors, again with applications to elasticity and hydrodynamics. In chapter 4, the algebra of matrices is developed and it is in this connexion that Rayleigh's principle is introduced. Here also we find the matrices of unitary field theory, and the reciprocal lattice used in X-ray analysis, with even a few words on integral equations.

The fifth chapter deals with multiple integrals, with of course Green's and Stokes's theorems. The problem of defining the area of a surface is presented properly, without the rash statements which appear in more elementary books.

Whereas each chapter so far mentioned deals with a subject which can be specified mathematically, without reference to applications, chapter 6 is headed *Potential Theory* and deals with the matter in a manner rather reminiscent of Routh's. It is immediately followed by a valuable chapter on operational methods. This is a subject on which the male author has already written, and the treatment here follows his earlier methods fairly closely; the fact that this method is not identical with the method using the Fourier-Mellin theorem is stressed. The next chapter, number 8, is on physical problems soluble by operational methods, including the design problems of seismographs.

One of the most instructive chapters in the book is that on *numerical methods*, where the classical interpolation formulae are not only derived, but a really careful discussion of the relative advantages of their different forms is given. Strategems for dealing with difficult regions are not overlooked, either. Here we find the Euler-Maclaurin theorem, and also methods of solving algebraic and differential equations numerically. Perhaps it would have been asking too much to beg for the inclusion of the relaxation method as well.

With chapter 10, we return to mathematical formulation, the subject being the calculus of variations. This is seen chiefly in its dynamical aspects, Hamilton's canonical equations appearing here. It might have been conducive to clear apprehension by the student if some consideration of the relation between the calculus of variations and the theory of integral equations could have been introduced into this chapter. The following chapter, one of the longer ones, is on functions of a complex variable. With the next chapter, which treats in detail of contour integration it has most of the material found in standard treatises on the subject, except the most advanced ones. This chapter on contour integration deals also with Bromwich's integral, needed in the Fourier-Mellin treatment of partial differential equations. The general subject of complex variable is continued again in the next chapter, which takes up the subject of conformal transformation, so important in physical applications, and so fundamental in Riemann's general theory. In this particular treatise, the emphasis is strongly on the physical applications, and one is glad to note that the Joukwosky aerofoil figures here. Incidentally, there is nowhere in this book any *geometrical* treatment of inversion, sources and sinks and their superposition. It is not suggested that this is an omission; it is rather a preference, which the reviewer shares, for analytical methods in practically every case. Chapter 14, on Fourier's theorem, differs markedly from any students' treatment known to the reviewer. Whilst the derivation of Fourier's theorem is rigorous, within the limits for which it is to be proved, it is as a whole written with more attention to applications than usual. The chapter contains Weierstrass's theorem on approximations by polynomials, and a paragraph on the detection of periodicities in experimental material.

Chapters 15 (Factorial, gamma and beta functions), 21 (Bessel functions), 23 (confluent hypergeometric functions), 24 (Legendre and associated Legendre functions) and 25 (elliptic functions) each deal with special functions of use in applied mathematics. In dealing with Bessel functions, there are some innovations of notation, whose value will be best judged when they have been considered for a few years. The confluent hypergeometric functions are naturally less used than the others mentioned, largely because of the difficulty of tabulation, but the Hh_n functions have now been tabulated, and receive their due share of attention. It is satisfactory to notice that, a general theory being less important than applicability, the Weierstrass form of the elliptic functions does not appear.

Interspersed between these chapters are others on particular mathematical topics. Thus, number 16 deals with the (general theory of) linear differential equations of the second order, and their solution by numerical methods, expansion of the solution in power series or in asymptotic expansions, and substitution of definite integrals. In chapter 17, on asymptotic expansions, there is a most useful treatment in which the method of steepest descent is fitted into perspective. A short chapter is devoted to the equations of mathematical physics, and others deal individually with wave propagation, diffusion of heat, and the applications of Bessel functions. In all these chapters, the modern methods, operational and by use of Bromwich integrals, are freely used.

From this long description, it will be seen that the book covers an enormous field, and that it has throughout a modern outlook, and puts physical applications into prominence. It is true, at the same time, that rigour is always regarded ; and there are several instances where theorems (or at least sets of conditions for the truth of theorems) are given for their physical use, which would be unfamiliar to most pure mathematicians. Each chapter is headed with a quotation, from standard literary works or elsewhere, and some of these show considerable humour. Chesterton's *Flying Inn*, for example, provides the couplet about the merry road which we did tread the night we went to Birmingham by way of Beachy Head, which is attached to the chapter on differential equations. Appended to each chapter is a selection of examples for practice. These are mostly from the Mathematical Tripos, and few of them could be called easy, but they certainly provide a means of testing the reader's grasp of the matter he has read.

Finally, there is an appendix of notes, and another containing a plea for a standardized, and new, notation for potential functions. Some will no doubt hail this gladly and others will be angered, but the present reviewer must confess to a lack of interest. The notation chosen should be clear on each occasion, but the advantages of a standard one seem to be small, more especially as we shall not cease to read papers in which other, and older, notations were used.

Many an advanced student will profit by the treatise, and so will the students of all advanced teachers who find the time to ponder the book and consider what lessons they may gain from it for use in their own teaching.

J. H. A.

Piezoelectricity, by W. G. Cady. Pp. 806. (New York: McGraw-Hill Book Co. Inc., 1946.) $9.00.

The Scott Laboratory, Wesleyan University, Middletown, Conn., is probably the most active centre of research on piezo-electricity in the world, and Professor W. G. Cady, who initiated work on that subject there, still leads and inspires it. That is equivalent to saying that no physicist is better equipped for producing a standard work on piezo-electricity than is Cady, whose treatise will earn him the gratitude of all interested in crystal physics.

The book starts with a relation of the discovery of pyro- and piezo-electricity, and of the parts played by the brothers Curie, by Kelvin and by Voigt in the study of phenomena and the development of theory. Short biographical notices of J. and P. Curie and W. Voigt are given. It is characteristic of Cady that this introduction, like the rest of the book, abounds in unselfish tributes to the work of others with hardly a mention of his own great contribution. A concise yet adequate introduction to crystallography is followed by chapters on crystal physics, including elastic and dielectric properties, and a treatment of thermodynamic potentials, which are used subsequently in the exposition of the theory of piezo-electricity. Valuable formulae and tables for transformation to rotated axes are given, and indeed, wealth of data and references is a special feature of the work. From a critical survey of all available experimental data, Cady has assigned " most probable values " to the elastic, dielectric, piezo-electric and other constants of quartz, tourmaline and Rochelle salt. This evaluation must indeed have been a laborious task, but it will prove of immense value to all workers in this field. Good line diagrams of idealized Rochelle salt and quartz crystals are given, but, a trifle oddly, none of tourmaline.

In the theory of piezo-electricity given by Voigt, piezo-electric strain is expressed in terms of applied electric field. Measurements show, however, that in the case of Rochelle salt the quotient of strain to field varies greatly with temperature, as also does the permittivity; whereas if the piezo-electric strain is expressed in terms of electric charge or electric displacement, the quotient is almost independent of temperature. Although Cady remarks that experimental results taken by themselves lead to the expression of piezo-electric strain in terms of displacement, " nevertheless, one should not confuse that which is most easily measurable with that which is most fundamental ; and if the proportionality of stress with field has to be abandoned, it appears fundamentally more logical to assume proportionality with polarization than with a parameter that involves both P and E". Since, however, it is always charge rather than polarization which is observed, might it not be held that expression in terms of charge is preferable ?

As was to be expected from the author, the chapter on the piezo-electric resonator bears the stamp of thoroughness. Cady has long been interested in the effect of air-gap on response frequency, to which he devotes a considerable proportion of this and subsequent chapters, but certain discrepancies between measured and predicted effects of gap still appear to defy explanation.

The reader may find that Chapter 14, which deals with circle diagram representation of the behaviour of piezo-electric resonators, requires more concentration than the rest of the work, with the exception of the treatment of the theories of Rochelle salt, but it does provide the nearest approach to visualization of the factors involved.

Methods of determining the axes of quartz crystals are described, more space being devoted to etching and optical than to X-ray methods. Chapter 17 deals briefly with various types of cut, with reduction of thermal coefficients of frequency and with vibration patterns, whilst Chapter 19 is a useful summary of knowledge on the piezo-electric valve-maintained oscillator, with clear, concise physical explanations of stabilizer and oscillator action.

Rochelle salt is a much more highly piezo-electric crystal than either quartz or tourmaline, but unlike these its dielectric and piezo-electric properties exhibit marked non-linearity and dependence on temperature. The Seignette electrics, of which Rochelle salt is the best known member, stand in relation to quartz or tourmaline pretty much as iron to paramagnetic substances, and Valasek, Kurchatov, Fowler, Mueller and Cady have worked on interaction, polarization and other theories in an endeavour to explain observed phenomena. Although about a quarter of the work under review deals with this subject, the impression left is that the explanations are still very imperfect ; but no doubt, as with ferromagnetism which is treated in an appendix, the reason lies in the complexity of the phenomena.

In this short review it is not possible to give an account of all the aspects of crystal physics or electrical engineering treated ; suffice it to mention the piezo-optic, electro-optic and other optical effects, the atomic structure of some piezo-electric crystals, diffraction of light by ultrasonic waves, all of which are briefly treated, and descriptions of methods of measurement with some hints on production of vibrators—not many, but enough to bring home the importance of the matter in a book which does not attempt an exhaustive treatment of engineering aspects.

Several chapters are followed by lists of references, and the bibliography at the end has over 650 entries. Mention is made again of the valuable data interspersed throughout the text because wealth of information is one of the principal features of the treatise ; for this reason, and because of the excellence of the general treatment of piezo-electricity, it will be in demand in every physics and engineering library. From the outset the reader is conscious of the intense enthusiasm of the author for his subject, and can hardly fail to admire the thoroughness and elegance of treatment. The printers are to be congratulated on the presentation, whilst the paucity of misprints bears witness to the meticulous care taken by all concerned. Finally, the apt quotations introducing each of the 31 chapters provide a proof —if indeed one were needed—that scientific ability and proficiency in classical literature are in no way mutually exclusive.

P. VIGOUREUX.

Applied Mathematics for Engineers and Physicists, by L. A. Pipes. Pp. xiii + 618. (New York and London: McGraw Hill Book Co. Inc., 1946.) 27*s*. 6*d*.

In the days of the giants—Maxwell, Kelvin, Rayleigh, J. J. Thomson—a physicist was a man who had learnt his mathematics first, and then turned to the application of mathematics to the material world, often taking up experimental work as well. If, then, they felt a lack of mathematical power to deal with a problem, it was because mathematics as a whole lacked the particular technique, and they were able to set about filling the gap, as Kelvin devised the method of inversion, and Rayleigh that of calculating the frequencies of a system, or as Maxwell systematized spherical harmonics and Thomson applied conformal transformation to electrostatic problems. Engineers learned through the experience of their predecessors, and, with a few exceptions, found need for little mathematics.

The position is now quite different. Physicists and engineers are trained in the laboratory, and acquire some mathematics in the earlier parts of their student careers, but usually have little time for definite study later. It seems to be quite common for them to feel the lack of adequate mathematical skill after a few years, and the reviewer can testify that many have confessed this to him. There were until recently only two ways in which this situation might be handled ; accepting it, and renouncing the hope of becoming mathematically skilled, or studying mathematics as a mathematician does, after entering on one's career as physicist or engineer. Demand, however, calls forth a supply, and books began to appear which were specially designed to fill the gap felt by practical scientists. I am not sure, but I suppose the treatises by Partington and by Mellor were among the earliest, if we except the writings of John Perry. They were designed for those who had little mathematics, and had to devote space to relatively simple integration and the like. As time went on, and science students had more mathematical background, these special textbooks could take up more advanced topics, such as convergence of sequences and integrals, or elementary matrix theory, especially when the needs of quantum mechanics brought the latter subject into prominence for physicists and even for chemists.

Books of this type, then, in which matters of mathematical technique are reviewed with special reference to the needs of scientists who are presumed to have some, but not sufficient, mathematical knowledge, have tended to increase in numbers. They have appeared in Britain and America, and in Germany, and have been written sometimes by those who are themselves noted as original workers (the Jeffreys in this country, Biot and v. Karman in U.S.A., for example) and sometimes by those who have given their efforts more completely to the problems of education (the Sokolnikoffs in U.S.A. or Houstoun in this country).

Among such books, the one by Professor Pipes of Harvard must take a very high place. The didactic skill of the author shines through the whole book, and his choice of topics is valuable indeed. After a chapter on infinite series, and one on the complex variable (not the functions of such a variable), he takes up the subject of Fourier series and then that of determinants and matrices. Each chapter is followed by a number of examples, not all of them easy, and by a list of references, which, even in these four chapters, includes the standard treatises of Goursat, Bromwich, Whittaker and Watson, and Bôcher (here spelt as Bôcker). The next topic is numerical solution of algebraic equations, and then comes the theory of linear differential equations ; here the Laplace transform is introduced, and there is a table of transforms containing 84 entries. In chapters 7, 8 and 9 the methods so far described are applied to electrical and mechanical problems. This gives the occasion for introducing the notion of normal co-ordinates (matrix methods being prominent here for transformation of variables) and the Rayleigh method of calculating natural frequencies.

In the chapter on finite differences, the theory is applied to electrical filters as well as to numerical integration. Later chapters, for which we have no space to give detailed descriptions, deal with partial differentiation and the calculus of variations, individual functions including those of Bessel and of Legendre, but not elliptic functions, and vector analysis.

Chapters 16 to 21 are devoted to the solution of the wave equation and the Laplace equation, with a chapter on complex functions interpolated at an appropriate point so that the use of conjugate functions for this purpose can be used, and with the operational method (which is not distinguished from the method of the Laplace transform) occupying the last of these chapters. The book closes with a much more difficult subject, that of non-linear oscillatory systems.

The contents have been set out in considerable detail, because whether a reader of the review desires to buy the book will depend on whether these are the subjects he wants to see treated. If he does, he may be assured that they are here well presented. J. H. A.

The Kinetic Theory of Liquids, by J. FRENKEL. Pp. xi + 485. (Oxford: The Clarendon Press, 1946.) 40*s*. net.

The formulation of a mathematical theory of the liquid state has attracted the attention of many workers, especially in the years just preceding the outbreak of war. It has proved to be a task of extraordinary difficulty. While the kinetic theory of gases and many aspects of the theory of the solid state have reached a highly developed mathematical form, the theory of liquids remains still, by comparison, largely empirical and descriptive. In the older theoretical considerations, emphasis was laid mainly on the analogy between the liquid state and the state of a highly compressed gas. This followed naturally from the success of theories of the kind usually associated with the name of van der Waals, which connect directly, along a single isotherm, the properties of the liquid and the gaseous states. More recently, as a result of the detailed study of the kinetics of lattice distortion in crystals, and perhaps also because of the insight gained in the study of the order-disorder transformation in alloys, it has become clear that there exists a close analogy between the liquid state and the state of an internally distorted crystalline solid. It is from this latter point of view that Professor Frenkel develops much of the theory included in the book under consideration.

To elucidate the theory the author has collected and correlated a good deal of detailed information concerning the condensed state of matter. This is woven together and unified with the theoretical work into a very readable and, indeed, fascinating book. The mathematics employed is throughout clear and simple—consisting, in fact, in many parts of the book, of simple algebraic equations relating such quantities as the activation energy, the times of relaxation, and various diffusion coefficients. It is a satisfactory feature of the book that the degree of elaboration of the mathematics is nicely related to the nature of the assumptions upon which particular calculations are based. Besides making the book easily readable, the elimination of over detailed mathematics avoids giving a false impression of the existing state of the theory.

It is possible here to mention specifically only a few of the many topics dealt with in this book. The opening two chapters are concerned with the kinetics of the crystalline state, and deal with such matters as hole formation, lattice distortion, diffusion, and the order-disorder problem. These chapters pave the way for a general discussion of the liquid state and the kinetics of fusion. Questions are raised such as: What is the immediate cause of melting? Why does melting constitute a transition of the first order, with a discontinuity in the entropy, rather than one of the second order such as occurs in the disordering of alloys? These and many other questions are critically examined in the light of the various theories put forward in recent years by the numerous workers in this field. A chapter of special interest deals, in the same critical spirit, with theoretical work on the electrical polarization of dipolar liquids, and also with the scattering of light by liquids composed of optically anisotropic molecules. There is a chapter on surface phenomena and, finally, one dealing with the properties of high polymeric substances and their solutions.

In spite of, or perhaps on account of, the clearly provisional character of the theoretical subject matter, this book is likely to prove valuable to a wide field of research workers, both experimentalists and theoreticians. Apart from the critical discussion of modern theories, the book is of great value for the many clear descriptions of relevant observational data only to be found elsewhere widely scattered throughout the literature, much of it in the Russian journals. One minor criticism should, perhaps, be made. The value of most scientific text-books, and certainly one of this kind, is greatly enhanced by a detailed index of the subject matter. It is a pity, therefore, that the present volume contains nothing more useful, in this respect, than a name index.

H. JONES.

Physical Science in Art and Industry, by E. G. RICHARDSON. Second Edition. Pp. xi + 299, with 77 illustrations. (London: The English Universities Press Ltd., 1946.) 15*s.* net.

In reviewing a second edition, it is customary to refer in detail to such alterations and additions to his earlier book as the author may have made, and to congratulate him upon the evident appreciation of his efforts. These things may well be taken for granted: the writer has enriched his pages, and is entitled to a warm welcome on that account.

The previous arrangement still holds; there are chapters dealing with locomotion, "out-of-door" physics, applications to the Fine Arts, and so forth. It is remarkable how stimulating it all becomes under Dr. Richardson's guidance. It is as if one was being taken a grand "tour du propriéteur" around the farm, along the river bank, down the mine, back via the Art Gallery—with a glimpse at contemporary architecture—finishing up in the kitchen. (This is not the exact order in the book, but all these items are there.) Now the point is this: these subjects reveal quite naturally how much applied physics have helped them both to progress and to conserve. It would be unreasonable to expect completeness in such a programme; what is noteworthy is how little of real importance is omitted.

Readableness is not secured by superficiality. The treatment of the hysteresis loop for soils (*p*H against moisture-content), and much the same for wood, are examples of how these quite advanced conceptions can be simply explained. To what extent they are fully grasped and used in the appropriate quarters is a different matter. Another instance is a brief account of air-conditioning, a subject upon which the wildest opinions are sometimes heard—the outcome of mental confusion about what the process is intended to do. Nobody could fail to understand the principles in this case. Like all efforts of a missionary nature, there is a finite risk that the reader will go on his way rejoicing, in the belief that he has been given the "know-how" and that all the rest is easy. He may soon discover his mistake, but he must not blame Dr. Richardson, whose skill in explanation may have been a trifle intoxicating. With this caution, it is more than likely that somebody may be truly inspired by working through this book carefully, and pondering upon its implications.

At the same time, the philosophical reaction to all this activity needs consideration, namely the impact of these applications upon physics itself. Probably the ever-increasing stress of circumstances, and the urge towards increased technical efficiency (especially in "Physics down the mine") will produce advances into the border-country between physics and engineering, and "sideways" along with physical chemistry.

If in fact there is no time like the present, then the appearance of an extended version at this moment of a first-class piece of work is an event to be welcomed, and used with a full sense of responsibility.

F. IAN G. RAWLINS.

The Theory of Functions of Real Variables, by LAWRENCE M. GRAVES. Pp. x + 300. (New York and London: McGraw Hill Book Co. Inc., 1946.) 20*s.*

The subject of "real variables" can mean anything from those parts of elementary calculus which can be dealt with intuitively up to the subject dealt with by Littlewood and by Hobson in their treatises; in general, one associates it with the more recondite theory of Fourier series, and feels that it has little of value (though maybe much of interest) for the applied mathematician.

This book, by the professor of mathematics at Chicago, sees the subject in a different light. The general view is that of the pure mathematicians named above, but it is written for those whom the author describes as "beginning graduate students", and considerable effort has been devoted to finding and including those results of interest to the applied mathematician.

The first chapter, on formal logic, is probably the most difficult in the book. It is followed by one on the number system, after which sets of points are considered, and then begins the "ordinary" mathematics—functions, differentiation, the Riemann integral, convergence and series, all treated by the methods of real-variable theory. In chapter 9, which deals with differential equations, are several useful theorems which show beforehand whether a differential equation will have a solution differentiable with respect to a parameter, and similar theorems. The last 125 pages are given to the general theory, mainly the properties of the Lebesgue and the Stieltjes integral, and form an excellent and relatively elementary introduction to the subject.

Whilst it would not be true to say that the book is likely to be of great direct value to the mathematical physicist, it is true that it is likely to be of great interest, and that it is well written and offers him a carefully planned survey of the present extent of its territory.

J. H. A.

Retinal Structure and Colour Vision. A restatement and an hypothesis, by E. N. Willmer with a Foreword by W. D. Wright. Pp. xii + 231. (Cambridge: The University Press, 1946.) 21*s*.

This book gives concisely and objectively a great deal of information on the structure and function of the retina. But its main purpose is to develop certain new ideas on the retinal end-organs and their rôle in vision. According to the main hypothesis, in its final form, stated in an Addendum, the end-organs in the human retina are of three types: cones, dark-adapting rods (absent in the central fovea) and non-adapting or so-called day-rods. As in the usual theory, scotopic vision is mediated by the dark-adapting rods. But the assumption of three kinds of cone is rejected and the three mechanisms demanded by the phenomena of colour vision are associated with end-organs in accordance with the following scheme: "red" mechanism, cones only; "green" mechanism, cones and day-rods linked to common bipolar cells and forming hybrid receptor units; "blue" mechanism, dark-adapting rods whose spectral sensitivity curve at the intensity levels required for colour discrimination has become displaced towards the blue. It may seem that the new view is in similar difficulties to the old in that it postulates structurally indistinguishable adapting and non-adapting rods outside the fovea and structurally indistinguishable cones and day-rods within the fovea. However, this is not the place to pursue criticisms of the actual hypothesis. The presentation of the hypothesis is marred by the attempt made to develop first a simpler theory involving only two colour mechanisms served respectively by rods and cones. This leads to difficult comparisons between colour diagrams based on what is essentially a theory of dichromatic vision, with standard colour diagrams referring to a trichromatic eye. Later in the book the third mechanism is introduced, but it is not clear to what extent the earlier arguments are regarded as still valid.

Despite these criticisms Dr. Willmer has many interesting things to say. In attempting a new interpretation of colour vision from a histological standpoint, he provides much food for thought for those who approach the subject with a physical background. W. S. S.

Tables of Fractional Powers, prepared by the Mathematical Tables Project. Pp. xxx + 486. (New York: Columbia University Press; Agents for U.K., Scientific Computing Service, Ltd., Bedford Square, W.C. 1, 1946.) $7.50.

Tables of x^n to cover a really considerable range in both variables would of course be impracticable, and the present volume does not attempt to deal in a comprehensive way with all the possibilities. The biggest single table in the book gives x^n when $1000x$ is any prime between 100 and 1000 and n ranges from 0·001 to 1. The next largest table gives the powers from 0·001 to 0·99 of the numbers $N/100$, where N ranges from 1 to 100, and this is supplemented by tables 1 and 2, covering the cases where x is an integer less than 11 and n is once more between 0 and 1.

There is also a table of π^n, where n is between 0 and 1, and, apart from a " miscellaneous" table, the remaining ones deal with fixed values of $n(\pm\frac{1}{2}, \pm\frac{1}{3}, \pm\frac{1}{4}, \pm\frac{2}{3}, \pm\frac{3}{4})$ and give values for $x=0$ to 10. Practically every table is to 15 decimals.

With these tables, and a calculating machine, it is possible to build up most of the results likely to be wanted. The tables where x is an integer can be combined with those where it is a fraction and those for n greater than 1 with those for smaller n. Moreover, by factorizing a number between 100 and 1000, any power of it may be formed as a product from the table which takes $1000x$ as a prime.

It is pointed out in the *Introduction*, and is perhaps worth noting, that the table of 10^n is in fact a table of antilogarithms, though it would not lend itself readily to use in logarithmic computation. Attached to the *Introduction* is a bibliography, the size of which is quite astonishing ; it contains 76 entries, of which about eight are powers of e, 14 are square or cube roots, and 22 are antilogarithms. Tables of $(1-r^2)^{\frac{1}{2}}$ are given in the bibliography, but not in the tables.

The tables are reproduced from typescript and the volume is well bound, but is printed on an unattractively dark paper.

J. H. A.

Conduction of Heat in Solids, by H. S. CARSLAW and J. C. JAEGER. Pp. vi+386. (Oxford: The University Press, 1947.) 30*s*. net.

In case others find the publications of Carslaw on classical mathematical physics confusing, it may be useful first to list them : his first publication was ***Fourier's Series and Integrals and the Mathematical Theory of Heat Conduction***, published in 1906. When this had been out of print for some time, he issued a revised edition in two volumes, of which the first, issued in 1921, was called ***Introduction to the Mathematical Theory of the Conduction of Heat in Solids*** ; the second volume, entitled ***Introduction to the Theory of Fourier's Series and Integrals***, appeared in the same year. In 1941, with Jaeger, he published ***Operational Methods in Applied Mathematics***, and now we have ***Conduction of Heat in Solids***, by the same two authors.

The oldest book dealt with the theory of the flow of heat governed by the equation of diffusion, and introduced the theory of Fourier's series very much as the French pioneer himself did. The methods based on conformal transformations received little attention, but the methods of images and of sources and sinks were well treated. Although Heaviside had carried out his pioneer work, ten years were to elapse before Bromwich gave it a sound mathematical basis. In the 1921 revision, the volume on heat conduction contained much the same material as the first one, but a chapter was added giving the solutions of problems which required the new methods ; the scheme preferred was that of Bromwich, expressing the solution as a contour integral and evaluating it by the method of residues. The theory of Fourier series was based on rigorous mathematical methods appropriate to the real variable—that is, by basing the work on the theory of measure and the Lebesgue integral.

The latest version has dropped all interest in the general theory of Fourier series and reverted to the outlook of the first version, where the theory of heat conduction takes priority. It differs from the forty-year-old work by including more about the methods based on complex variables, both in the guise of transformation theory and in operational methods. In this respect, it is complementary to the other book written jointly by these two authors. Here, like most recent writers, they prefer the method (not really operational) based on the Fourier-Mellin theorem, and they give a table of functions and their transforms. The book has an appreciably more practical outlook than had the older editions. There is less interest in the flow of heat in bodies of odd shapes, or heated in unlikely ways, and more in problems where the bodies and boundary conditions can be approached in practice. One boon which follows from this more practical outlook is the addition of tables of the error function and its derivatives and integrals, and of the roots of $x \tan x-C=0$, $x \cot x+C=0$, $x\, J_1(x)/J_0(x)=C$ and $J_0(x)\,Y_0(Cx)=Y_0(x)J_0(Cx)$, each for a number of values of C.

In its old forms, the book has been a standard work for as long as the reviewer can remember. The new form is likely to serve the needs of another generation at least as well.

J. H. A.

Tables of Spherical Bessel Functions, by the Mathematical Tables Project, National Bureau of Standards, Volume 1. Pp. xxviii + 375. (New York: Columbia University Press.) $7.50.

The functions tabulated in this volume are not, strictly, Bessel functions, but the quantities $J_n(x)/x^{\frac{1}{2}}$ with a normalizing factor. They are thus the functions which are directly required in solving the wave equation in spherical, conical or spheroidal co-ordinates, for they are given for values of n equal to $\pm(N+\frac{1}{2})$ where N is an integer from 0 to 13 inclusive. As Dr. Morse says in the Foreword: "It seems a far cry from the diffusion of a search-light beam by fog to the triggering of nuclear disintegrations by neutron collisions ; from the production of "knock" in gasoline engine cylinders to the electrical oscillations of an ultra-high-frequency radio tube"; but all these phenomena depend on the wave equation, and consequently may call for these functions.

The tables are given to at least eight significant figures for x up to 10, where the interval is 0·01, and to seven figures from $x=10$ to $x=25$ (interval 0·1). In general, the functions with their (modified) differences are given, but near the zeros of the functions these would not suffice for interpolation to full accuracy, and then the product of the function and a suitable power of x is tabulated in a subsidiary table so as to remove the irregularity.

The Introduction, as usual, contains much interesting matter, including series which relate these functions to the sine and cosine integrals. J. H. A.

Introduction to Atomic Physics, by S. TOLANSKY. Second Edition, with Appendix. Pp. 351. (London: Longmans, Green and Co.) 15*s*.

In Dr. Tolansky's book the student is presented with a survey of the whole of modern physics in 350 pages. The treatment uses only the simplest mathematics, but nearly all the traditional derivations of fundamental formulae and a few new ones have been worked into the text. On the whole the treatment is too concentrated. A student who is to understand the subject matter of the book will have to bring a very good knowledge of classical physics to bear before the full implications of many of Dr. Tolansky's concise phrases are apparent to him. However, as a guide to a student who wishes to select subjects for extensive study the text is excellent. Some references to standard texts are made at the end of each chapter, and these could with advantage be increased in a further edition. On the whole the book is free from misstatements of fact, and such as there are can fairly be attributed to an attempt to gloss over difficulties. Few would agree that the electrolytic concentration of deuterium can be wholly accounted for by differences in mobility, and the historical account of the hydrogen spectrum is less than just to Rydberg's contribution to spectroscopy : the account of the determination of the range of α particles is unnecessarily old-fashioned. In speaking of nuclear isomerism, the statement that the iosmers can sometimes be separated chemically, although true, is misleading, unless more fully explained. The book concludes with a good elementary chapter on relativity, and a useful appendix on recent determinations of the atomic constants ; a second appendix summarizes the main part of the new scientific information contained in the Smyth Report. The binding and paper are reasonable and the price moderate. C. H. C.

The Vector Operator j, by F. C. GILL, A.M.I.E.E., A.I.Mech.E. Pp. 61, with 32 diagrams. (Pitman, 1946.) 7*s*. 6*d*. net.

This little book is written to explain to the mathematically uninitiated how the complex operator is used in the calculation of quantities in problems of alternating current engineering relating to single-phase, two-phase and three-phase circuits. Naturally no fundamental treatment is possible, nor is it attempted, within the limits of size adopted. The method demands care in choice of symbols and in their correct use. This correctness is usually attained in the examples selected, but a few slips occur which may cause confusion, and should be eliminated in any further edition. The book should be found helpful to engineers in the field indicated. D. O.